

Estimación de heredabilidades y correlaciones genéticas  
para caracteres funcionales mediante análisis de supervivencia  
en Holstein Colombiano.

*Tesis presentada para optar al título de Magister de la Universidad de Buenos Aires,  
Área Biometría y Mejoramiento*

**Yeni Liliana Bernal Rubio**

Zootecnista - Universidad Nacional de Colombia - 2006

Lugar de trabajo: Cátedra Mejoramiento genético animal. Facultad de Agronomía.  
Universidad de Buenos Aires.



Escuela para Graduados Ing. Agr. Alberto Soriano  
Facultad de Agronomía – Universidad de Buenos Aires

## **COMITÉ CONSEJERO**

### **Director de tesis**

Rodolfo Juan Carlos Cantet  
Ingeniero Agrónomo (Universidad de Buenos Aires)  
MSc (Montana State University)  
MSc (University of Illinois)  
PhD (University of Illinois)

### **Consejero de Estudios**

Luis A. Escobar  
Ingeniero Administrativo (Universidad Nacional de Colombia)  
MS (Inter-American Statistical Training Center)  
PhD (Iowa State University)

## **JURADO DE TESIS**

### **Director de tesis**

Rodolfo Juan Carlos Cantet  
Ingeniero Agrónomo (Universidad de Buenos Aires)  
MSc (Montana State University)  
MSc (University of Illinois)  
PhD (University of Illinois)

### **JURADO**

Daniel Omar Maizón  
Médico Veterinario (Universidad de Buenos Aires)  
MSc (Universidad de Buenos Aires)  
PhD (Cornell University)

### **JURADO**

César Gabriel López  
Ingeniero Agrónomo (Universidad Nacional de Lomas de Zamora)  
MSc (Universidad Nacional de Rosario)  
PhD (Oregon State University)

Fecha de defensa de la tesis: 16 de Diciembre de 2013

## Agradecimientos

En primer lugar, agradezco a mi familia por el apoyo incondicional que siempre me han brindado. Sus sabias palabras y oportunos consejos han sido de gran ayuda y motivación en todo momento. Llevar a buen término tanto esta investigación como los demás logros que he obtenido no hubiera sido posible sin su compañía.

Asimismo, quiero agradecer a Fito Cantet, que más que director de tesis ha sido un guía y maestro tanto en el aspecto personal como profesional. Muchas gracias por la paciencia, la confianza y por enseñarme el valor de trabajar de manera independiente y de ver todo con ojos críticos, siempre desde lo constructivo.

También quiero agradecer a mis amigos y compañeros de oficina: Andrés, Camilo, Carolina, José Luis, Juan David, María José, Natalia y Sebastián, así como a las docentes de la cátedra: Ana, Laura, Mónica y Valeria. Las agradables charlas acompañadas de café, humor y también opiniones y buenos consejos hicieron que me sintiera como en casa. No obstante agradezco inmensamente a Sebastián, quien dedicó muchas horas de su tiempo para ayudarme en esta investigación. Eso es algo inolvidable.

Igualmente he de agradecer al Dr. Daniel Maizón y al Dr. Vincent Ducrocq por sus valiosas opiniones y sugerencias en etapas claves de la investigación. También, a las instituciones que hicieron posible el desarrollo de esta tesis, desde el aspecto financiero (Consejo Nacional de Investigaciones Científicas – CONICET y Departamento Administrativo de Ciencia, Tecnología e Investigación - COLCIENCIAS) hasta el suministro de la información (Asociación Holstein de Colombia).

Finalmente, a todas y cada una de las personas con quienes compartí y me enseñaron muchas cosas valiosas a nivel personal y profesional.

**Declaración**

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en ésta u otra institución.

**Publicaciones derivadas de la tesis**

Bernal Rubio, Y.L. y Cantet, R.J.C. 2012. Efecto de caracteres de tipo en la vida productiva de vacas Normando colombiano. *Actas Iberoamericanas de Conservación Animal-AICA*. Vol. 2, p.347 – 354. *ISSN*: 2253-7325. España.

## INDICE GENERAL

<b>CAPÍTULO 1: Introducción .....</b>	<b>1</b>
<b>CAPÍTULO 2: Revisión bibliográfica .....</b>	<b>3</b>
<b>2.1. Definiciones y estado del arte para caracteres funcionales y reproductivos .....</b>	<b>4</b>
2.1.1. Caracteres reproductivos en la vaca lechera .....	4
2.1.1.1. Intervalo Primer-Último Servicio .....	5
2.1.1.2. Días Abiertos .....	6
2.1.1.3. Importancia económica de los caracteres reproductivos .....	7
2.1.2. Vida productiva en ganado lechero .....	8
2.1.2.1. Importancia económica .....	9
2.1.3. Sistemas de producción de leche en Colombia .....	9
<b>2.2. Metodología del Análisis de supervivencia .....</b>	<b>10</b>
2.2.1. Antecedentes del análisis de supervivencia en el mejoramiento genético animal .....	11
2.2.2. Definiciones generales .....	13
2.2.2.1. Censura .....	13
2.2.2.2. Descripción de los datos de supervivencia .....	14
2.2.3. Distribuciones de los tiempos de supervivencia .....	15
2.2.4. Métodos de estimación .....	16
2.2.4.1. Estimación bajo Modelos paramétricos .....	16
2.2.4.2. Estimación bajo Modelos no paramétricos .....	17
2.2.4.2.1. Estimador del producto límite o Kaplan Meier .....	17
2.2.4.2.2. Test de hipótesis bajo estimación no paramétrica .....	18
2.2.5. Modelos de regresión .....	19
2.2.5.1. Modelos de tiempo de falla acelerada .....	19
2.2.5.2. Modelos de riesgo proporcional .....	19
2.2.5.3. Estimación paramétrica en modelos de regresión .....	20
2.2.5.4. Estimación semi-paramétrica: modelo de Cox .....	21
2.2.6. Generalizaciones .....	21
2.2.6.1. Covariables tiempo – dependientes .....	21
2.2.6.2. Estratificación .....	22
2.2.7. Modelos de fragilidad .....	22
2.2.8. Tiempos de falla discretos – datos agrupados .....	24
2.2.9. Otros aspectos relevantes al análisis de supervivencia .....	24
2.2.9.1. Chequeo del supuesto de censura aleatoria .....	24
2.2.9.2. Bondad ajuste modelos paramétricos .....	25
<b>CAPÍTULO 3: Materiales y metodología .....</b>	<b>26</b>
<b>3.1. Base de datos .....</b>	<b>27</b>
3.1.1. Vida productiva (VP) .....	27
3.1.2. Intervalo primer-último servicio (IPU) .....	28
3.1.3. Días abiertos (DA) .....	29

<b>3.2. Metodología</b>	<b>30</b>
<b>3.2.1. Análisis de supervivencia</b>	<b>30</b>
3.2.1.1. Supuesto de proporcionalidad	30
3.2.1.2. Distribución Weibull para la función de riesgo base	31
3.2.1.3. Significancia de efectos fijos - Prueba del cociente de verosimilitud (Kalbfleisch y Prentice, 1980)	32
3.2.1.4. Extensión del modelo para incluir efectos aleatorios	33
<b>3.2.2. Modelo de datos agrupados</b>	<b>33</b>
3.2.2.1. Reparametrización del modelo de datos agrupados	35
3.2.2.2. Inclusión término frágil modelo de datos agrupados	35
<b>CAPÍTULO 4: Resultados</b>	<b>36</b>
<b>4.1. Vida Productiva</b>	<b>37</b>
4.1.1. Control del supuesto de proporcionalidad en las funciones de riesgo	37
4.1.2. Control de la factibilidad del ajuste Weibull para la función de riesgo base	38
4.1.3. Definición del modelo de efectos fijos	38
4.1.4. Estimación de componentes de varianza y heredabilidad	42
4.1.5. Tendencia genética	43
4.1.6. Correlación genética para vida productiva y producción de leche	43
<b>4.2. Intervalo primer-último servicio</b>	<b>44</b>
4.2.1. Modelo de datos agrupados	44
4.2.2. Estimación de componentes de varianza y heredabilidad	48
4.2.3. Tendencia genética	49
<b>4.3. Días abiertos</b>	<b>49</b>
4.3.1. Modelo de datos agrupados	50
4.3.2. Estimación de los componentes de varianza y la heredabilidad	54
4.3.3. Tendencia genética	55
4.3.4. Correlación genética para días abiertos y producción de leche	55
<b>CAPÍTULO 5: Discusión</b>	<b>57</b>
<b>5.1. Vida Productiva</b>	<b>58</b>
5.1.1. Ajuste Weibull para la función de riesgo base	58
5.1.2. Modelo de efectos fijos	59
5.1.3. Componentes de varianza y heredabilidad	60
5.1.4. Tendencia genética	61
<b>5.2. Intervalo primer-último servicio</b>	<b>62</b>
5.2.1. Modelo de datos agrupados	63
5.2.2. Estimación de componentes de varianza y heredabilidad	63
5.2.3. Tendencia genética	64
<b>5.3. Días abiertos</b>	<b>64</b>
5.3.1. Modelo de datos agrupados	65
5.3.2. Estimación de los componentes de varianza y la heredabilidad	66
5.3.3. Tendencia genética	66

<b>CAPÍTULO 6: Conclusiones .....</b>	<b>68</b>
<b>6.1. Vida productiva .....</b>	<b>69</b>
<b>6.2. Intervalo entre primer y último servicio .....</b>	<b>69</b>
<b>6.3. Días abiertos .....</b>	<b>69</b>
<b>BIBLIOGRAFIA .....</b>	<b>70</b>



## ÍNDICE DE CUADROS

<b>Tabla 3.1.1.</b> Medias y desviaciones estándar para diferentes variables productivas y reproductivas .....	27
<b>Tabla 3.1.1.1.</b> Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter vida productiva .....	28
<b>Tabla 3.1.2.1.</b> Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter intervalo entre el primer y último servicio .....	29
<b>Tabla 3.1.3.1.</b> Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter días abiertos .....	29
<b>Tabla 4.1.3.1.</b> Test de cociente de verosimilitud: factores incluidos secuencialmente ....	39
<b>Tabla 4.1.3.2.</b> Test del cociente de verosimilitud comparando el modelo completo vs. modelos excluyendo un efecto a la vez .....	39
<b>Tabla 4.1.3.3.</b> Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), tasa de riesgo (TR), y número de observaciones no censuradas ( $n_{noc}$ ), para cada efecto incluido en el análisis de vida productiva .....	40
<b>Tabla 4.2.1.1.</b> Test del cociente de verosimilitud, incluyendo factores secuencialmente al modelo de efectos fijos .....	45
<b>Tabla 4.2.1.2.</b> Test del cociente de verosimilitud comparando el modelo completo contra aquellos excluyendo de un efecto a la vez .....	46
<b>Tabla 4.2.1.3.</b> Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), Tasa de riesgo (TR), y número de observaciones no censuradas ( $n_{nocens}$ ), para cada efecto incluido en el análisis del carácter IPU .....	46
<b>Tabla 4.2.1.4.</b> Estimación discreta de la función base de supervivencia para el intervalo primer-último servicio .....	48
<b>Tabla 4.3.1.1.</b> Test del cociente de verosimilitud, incluyendo factores de forma secuencial.....	51
<b>Tabla 4.3.1.2.</b> Test del cociente de verosimilitud comparando el modelo completo vs. modelos excluyendo un efecto a la vez .....	51
<b>Tabla 4.3.1.3.</b> Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), Tasa de riesgo (TR) y número de observaciones no censuradas ( $n_{nocens}$ ), para cada nivel de los efectos incluidos en el análisis de días abiertos .....	52
<b>Tabla 4.3.1.4.</b> Estimación discreta de la función base de supervivencia para el carácter días abiertos .....	54

## INDICE DE FIGURAS

<b>Figura 2.2.1.1.</b> Distribución de frecuencias para vida productiva (días) en Holstein Colombiano .....	11
<b>Figura 3.2.1.1.1.</b> Ilustración del supuesto de riesgos proporcionales .....	31
<b>Figura 3.2.1.2.1.</b> Test gráfico para la distribución Weibull .....	32
<b>Figura 4.1.1.</b> Función de supervivencia $S(t)$ para vida productiva para los diferentes niveles de edad al primer parto .....	37
<b>Figura 4.1.1.1.</b> Supuesto de riesgo proporcional para vida productiva (VP) .....	37
<b>Figura 4.1.2.1.</b> Control del supuesto Weibull mediante regresión lineal de $\text{Log}[-\log(S_{km}(t))]$ versus $\log(\text{tiempo})$ para vida productiva .....	38
<b>Figura 4.1.3.1.</b> Tasa de riesgo para número de partos (NP), nivel de producción de leche (PL), efecto días post-primer parto (DPP), y edad al primer parto (EPP) .....	42
<b>Figura 4.1.5.1</b> Tendencia genética para vida productiva .....	43
<b>Figura 4.2.1.</b> Distribución de frecuencias para el carácter intervalo primer-último servicio .....	44
<b>Figura 4.2.1.1.</b> Tasa de riesgo relativo para número de partos (NP) .....	47
<b>Figura 4.2.1.2.</b> Estimación discreta de la función base de supervivencia para el carácter intervalo primer-último servicio .....	47
<b>Figura 4.2.3.1.</b> Tendencia genética para intervalo primer-último servicio .....	49
<b>Figura 4.3.1.</b> Distribución de frecuencias para el carácter días abiertos .....	50
<b>Figura 4.3.1.1.</b> Tasa de riesgo para edad al primer parto (EPP) y número de partos (NP) .....	53
<b>Figura 4.3.1.2.</b> Estimación discreta de la función base de supervivencia para días abiertos .....	53
<b>Figura 4.3.3.1.</b> Tendencia genética para días abiertos .....	55
<b>Figura 5.1.4.1.</b> Tendencia genética para vida productiva y producción de leche .....	62

## Abreviaturas

$C_i$  =  $i$ -ésimo tiempo de censura.

DA = Días abiertos.

DPP = Días post-primer parto.

EM = Algoritmo Esperanza – maximización (Dempster et al., 1977).

EPP = Edad al primer parto.

$f(t)$  = Función de densidad.

$F(t)$  = Función de distribución acumulada.

FP = Fecha de parto.

$h^2$  = heredabilidad.

$h(t)$  = Función de riesgo.

$H(t)$  = Función de riesgo acumulada.

IEP = Intervalo entre partos.

IPU = Intervalo entre primer y último servicio.

$\log(S(t))$  = logaritmo de la función de supervivencia.

MRP = Modelos de riesgo proporcional.

NP = Número de partos.

PL = Nivel de producción de leche.

RR = Riesgo relativo.

$r_s$  = Correlación genética simple.

$S(t)$  = Función de supervivencia.

$S_0(t)$  = Función base de supervivencia.

$S_{km}(t)$  = Estimador Kaplan-Meier de la función de supervivencia.

SA. = Análisis de supervivencia.

SC = Servicio por ciclo.

$T_i$  =  $i$ -ésimo tiempo de supervivencia.

TR = Tasa de riesgo.

TU = Variable Time-Unit generada por Survival Kit V6.0 (Ducrocq et al., 2010).

VP = Vida productiva.

$\delta_i$  =  $i$ -ésimo indicador de censura.

$\sigma_G^2$  = Varianza genética.

$\sigma_{hy}^2$  = Varianza de hato-año.

$\sigma_a^2$  = Varianza aditiva.

## RESUMEN

Las evaluaciones genéticas para caracteres funcionales requieren metodologías como el análisis de supervivencia (SA), que contemplen registros pertenecientes a hembras que no han presentado el evento de interés (descarte o preñez) al momento de la evaluación. Así, los objetivos de esta tesis fueron: 1) analizar los factores que afectan el riesgo de descarte en vida productiva (VP), y de concepción en días abiertos (DA) e intervalo primer-último servicio (IPU), 2) estimar parámetros genéticos de dispersión para los caracteres evaluados. Se emplearon 44652 registros de lactancia de 15706 vacas Holstein Colombiano, ajustando un modelo animal frágil Weibull para VP, con el número de partos (NP), producción de leche (PL) y edad al primer parto (EPP) como efectos fijos. En el caso de DA e IPU se emplearon 14789 servicios de 6205 vacas, ajustando un modelo de datos agrupados con efectos fijos de NP, EPP y número de servicios (SC). En todos los casos se incluyeron efectos aleatorios de animal y de hato. Tanto PL como NP tuvieron gran influencia en la VP, al igual que NP en DA e IPU. Se estimaron valores de  $h^2$  de 0,104, 0,086 y 0,1013 para VP, DA, e IPU, respectivamente, así como correlaciones genéticas simples ( $r_s$ ) entre el nivel de PL con VP y DA de  $-0,03$  y  $0,47$ , respectivamente. Teniendo en cuenta tanto los valores de cría expresados como riesgo relativo, así como la magnitud de  $r_s$ , un aumento en PL conllevaría a disminución en el riesgo de descarte y aumento en el riesgo de concepción, dando lugar a aumentos en la VP y reducción en DA. Las  $h^2$ s estimadas sugieren la existencia de variabilidad genética para VP, DA e IPU en Holstein Colombiano, sirviendo de sustento para la implementación de evaluaciones genética de la raza, aprovechando las ventajas de metodologías como SA.

## ABSTRACT

Genetic evaluations for functional traits require using methodologies, as survival analysis (SA), that take into account records from females which have not presented the event of interest (culling or pregnancy) at the end of analysis. The objectives of this research were: 1) to evaluate the factors affecting culling risk for productive life (VP) and conception risk for open days (DA) and interval from first to last insemination (IPU); 2) to estimate genetic dispersion parameters for those traits. Records on 44652 lactations from 15706 Colombian Holstein dairy cows were used for analyzing VP, adjusting a Weibull animal frailty model with parity (NP), milk yield (PL) and age at first calving (EPP) as fixed effects. For DA and IPU, 14789 insemination records from 6205 cows were used, using a grouped data model, with NP, EPP and number of inseminations by estrous cycle (SC) as fixed effects. In all cases, an animal and herd effects were included as random effects. Both PL level and NP significantly affected VP, as well as NP for DA and IPU. Heritability ( $h^2$ ) estimates for VP, DA and IPU were 0,104, 0,086 y 0,1013, respectively, as well as simple genetic correlations ( $r_s$ ) between PL and VP, and PL and DA, were equal to -0,03 y 0,47, respectively. Because of breeding values were expressed as relative risks, the  $r_s$  values suggest that an increase in PL would be associated to a decrease in the culling risk, and to an increase in conception risk, leading to an increment in VP and to a reduction in DA. Values of  $h^2$  suggest the existence of genetic variability for VP, DA and IPU, giving support to genetic evaluations in Colombian Holstein, taking advantage of methodologies as SA.

**Palabras clave:** Holstein Colombiano; Análisis de supervivencia; modelos frágiles; modelo de datos agrupados; vida productiva; días abiertos; intervalo entre el primer y el último servicio; heredabilidad.

**Key words:** Colombian Holstein; Survival analysis; frailty models; grouped data model; productive life; open days; interval from first to last insemination; heritability.

## ***Capítulo 1. Introducción***

## **Capítulo 1. Introducción**

Debido a los cambios en las preferencias de los mercados, los objetivos de selección genética en los sistemas de producción animal se han modificado en relación con el pasado. Actualmente se centran en maximizar la eficiencia de los sistemas y la calidad de los productos obtenidos, más que la producción por sí misma. Surge así el concepto de “caracteres funcionales”, de modo de tener en cuenta tanto la eficiencia por parte del animal, así como la rentabilidad y el impacto que puedan generar dichos caracteres en los sistemas productivos. Tal es el caso de *vida productiva* (VP), carácter que se ha convertido en objetivo de las evaluaciones genéticas, debido a su efecto en la producción y en los ingresos del hato. Sin embargo, cuando aumentan los niveles de producción en los sistemas ganaderos, se afectan otras características funcionales como la fertilidad (Oltenacu y Broom, 2010). Esto es en parte, consecuencia de que la mayoría de los problemas reproductivos de las hembras se relacionan con problemas uterinos, ováricos o, en vacas con un bajo nivel nutricional, estrés metabólico asociado con el nivel de producción de leche (Bell y Roberts, 2007; Dobson et al., 2007; Sheldon et al., 2008). De este modo, caracteres como los *días abiertos* (DA), intervalo entre partos o el *intervalo entre el primer y el último servicio* (IPU) han sido incluidos en índices económicos debido a su efecto en la productividad.

Los análisis estadísticos de caracteres como VP, DA o IPU, son realizados empleando dos metodologías diferentes: los modelos lineales mixtos (Henderson, 1984), o el análisis de supervivencia. Uno de los primeros antecedentes de la aplicación de esta última metodología en la evaluación genética animal es el trabajo realizado por Famula (1981), quien propuso emplear análisis de supervivencia para analizar VP en ganado lechero. Posteriormente, Ducrocq et al., (1988a,b) emplearon modelos de supervivencia para VP, sentando precedente en el uso de esta metodología en el contexto del mejoramiento genético animal. En consecuencia, y teniendo en cuenta los antecedentes de la metodología del análisis de supervivencia en la evaluación genética animal, el presente trabajo tiene como objetivo principal utilizar y extender la metodología del análisis de supervivencia a la evaluación de los caracteres funcionales VP, DA e IPU en la población vacuna Holstein de Colombia. En primer lugar, se presentan los antecedentes y se realiza una revisión de la literatura sobre el análisis de supervivencia, así como las definiciones de cada uno de los caracteres a evaluar (Capítulo 2). En el capítulo 3 se describe la base de datos a analizar, así como las variables tenidas en cuenta al momento de organizar la información. Posteriormente, se desarrolla la metodología relacionada con el análisis de supervivencia y con el modelo de datos agrupados, empleados en la evaluación de los caracteres de interés (Capítulo 4). Los resultados y la discusión general se presentan en los capítulos 5 y 6, respectivamente. Finalmente, se desarrollan las conclusiones extraídas de la presente investigación.



## ***Capítulo 2. Revisión bibliográfica***

## **Capítulo 2. Revisión bibliográfica.**

### **2.1 Definiciones y estado del arte para caracteres funcionales y reproductivos.**

La evaluación genética animal es una herramienta de mejoramiento esencial para detectar individuos que sobresalen en los diferentes caracteres de interés económico, ya sean productivos, reproductivos, o de calidad de producto. En el ganado lechero, los caracteres reproductivos se han convertido en objetivo de programas de mejoramiento al ser enmarcados dentro de la categoría de caracteres funcionales. Su importancia reside en aumentar los ingresos del productor reduciendo los costos de inversión (Groen et al., 1997). Ejemplo de caracteres funcionales son la vida productiva, la facilidad de parto, la eficiencia alimenticia, la resistencia a enfermedades, y los caracteres reproductivos como la edad al primer parto, el intervalo entre partos, los días “abiertos” (como vaca no preñada) y la tasa de preñez, entre otros.

En primer lugar, la *vida productiva* ha sido uno de los caracteres de mayor importancia económica debido a su efecto sobre los costos de inseminación, en el nivel de producción diaria, y en el costo de reemplazo de hembras. No obstante, ha sido un carácter difícil de medir dada la presencia de observaciones provenientes de animales que continúan en producción al momento del análisis (observaciones censuradas). En cuanto a caracteres de fertilidad, los más utilizados para evaluar la eficiencia de programas de manejo reproductivo son los *días abiertos* y los indicadores de la tasa de concepción como el *intervalo entre el primero y el último servicio*. En el caso de los días abiertos, un incremento en dicho período implica pérdidas de ingresos relacionados con mayor número de días en período seco, así como menos terneros nacidos por año. Esta variable indicadora está compuesta, principalmente, por el período voluntario de espera, lapso necesario para que aparezca un primer celo después del parto y que responde tanto a variables fisiológicas como a decisiones del productor. Así, un incremento en el número de días abiertos estará relacionado con fallas en la detección de celos y en la concepción. De igual modo, resulta de interés evaluar el intervalo primer-último servicio, dado que un aumento en el número de días entre el primer y último servicio, estará relacionado con una baja tasa de concepción (Jorjani, 2005). Sin embargo, y con el fin de abordar de forma más adecuada la temática de estudio, se describirán en primer lugar los antecedentes y definiciones relacionadas con los caracteres de interés, y posteriormente, el desarrollo y fundamento teórico del análisis de supervivencia, metodología central del presente trabajo.

#### **2.1.1. Caracteres reproductivos en la vaca lechera.**

La fertilidad es el carácter que mayor impacta la rentabilidad de las empresas ganaderas (Dziuk y Bellows, 1983). En términos de la escala de medición, distribución en la población y parámetros involucrados, los caracteres de fertilidad pueden clasificarse en binarios, categóricos, continuos o normales, umbrales, y de supervivencia, entre otros. De particular interés económico son aquellos caracteres relacionados con la vida productiva de las hembras. A tal efecto, Bourdon y Brinks (1983) compararon caracteres como el intervalo entre partos (IEP), la edad al primer parto (EPP) y la fecha de parto (FP) como medidas de eficiencia reproductiva, observando que la selección por disminución de IEP podía tener resultados sesgados. Esto se debe a que IEP más cortos corresponden a vacas

cuyo primer parto ocurrió tardíamente, razón por la cual seleccionar dichos individuos puede resultar en selección indirecta por una edad a la pubertad tardía. De esta forma, los autores sugirieron tener en cuenta FP al momento de evaluar el desempeño reproductivo.

Por su parte, Meyer et al. (1990) observó que los valores de cría para caracteres de fertilidad en hembras son difíciles de estimar porque la expresión del potencial reproductivo es afectada por las decisiones del productor. Aun cuando las estimaciones de heredabilidad para fertilidad son menores a 0,05 (ver, por ejemplo, Freeman, 1984), existe cierta variabilidad genética que puede ser aprovechada (Jansen, 1985). Al respecto, el autor encontró como los principales factores que afectaron a la fertilidad el número de partos, el toro usado en el servicio, el inseminador y el efecto hato-año-estación. Mark (2004) puntualizó que la fertilidad de la hembra es un carácter complejo cuyo estudio puede realizarse empleando diferentes indicadores de desempeño reproductivo. Dicha afirmación es concordante con lo expuesto por Jamrozik et al. (2012), para quienes la fertilidad no puede ser totalmente representada por un único fenotipo.

Así, los caracteres de fertilidad que pueden considerarse para ser evaluados genéticamente son: la tasa de concepción, el número de inseminaciones necesarias para obtener una preñez, intervalo parto-concepción (días abiertos), intervalo primer-último servicio (o concepción) e intervalo entre partos consecutivos. En el caso de intervalo entre partos, este puede ser calculado a partir de datos relacionados con caracteres de producción de leche (fechas de parto). Sin embargo, los caracteres restantes requieren registros de inseminación o diagnóstico de preñez, datos que no son informados habitualmente debido a la carencia de un programa de registro adecuado (González-Recio et al., 2005).

#### **2.1.1.1. Intervalo Primer-Último Servicio**

Como se mencionó previamente, el intervalo primer-último servicio está relacionado con el tiempo desde el primer servicio, que tarda la vaca en alcanzar la preñez. Por lo tanto, en rodeos que emplean la inseminación artificial, es un indicador de la eficiencia de detección de celos y problemas relacionados (Warren, 1984). Así, hay animales que quedan preñados en el primer servicio o que necesitan dos o más servicios para lograr una preñez.

No obstante y de acuerdo con Swalve et al. (1992), existe además una dificultad estadística para describir adecuadamente los datos de fertilidad debido principalmente a su distribución sesgada. Sin embargo, empleando mínimos cuadrados los autores reportaron que el intervalo entre primer-último servicio se incrementó en relación con el aumento del nivel promedio de producción de leche dentro del hato, y disminuyó a medida que las vacas tuvieron un mayor número de partos. Este resultado, más evidente entre el primer y segundo parto con respecto a lactancias posteriores, puede ser explicado debido al fuerte descarte que experimentan las hembras después de la segunda lactancia. No obstante, dicho intervalo disminuyó a medida que el número de días al primer servicio postparto fue mayor. La heredabilidad del intervalo fue reportada entre 0,004 y 0,008, valores que son mayores que los estimados para caracteres como la tasa de no retorno al celo (Swalve et al., 1992).

Posteriormente, e incluyendo caracteres que pueden ser evaluados empleando fechas de servicios, Hodel et al. (1995) estimaron parámetros genéticos para diferentes medidas de

desempeño reproductivo, entre ellas el intervalo primer-último servicio, para el que informaron una heredabilidad de 0,038, valor que fue acorde con otros estudios (Janson, 1980; Weller y Ron, 1992). Además, encontraron correlaciones genéticas cercanas a  $-0,88$  entre la tasa de no retorno al celo y el intervalo primer-último servicio (IPU), valores que concuerdan con los reportados por Everett et al. (1966) y Hermas et al. (1987). Sin embargo, los autores aclararon que las correlaciones estimadas para IPU son sesgadas debido a la presencia de vacas de alta producción que tienen más oportunidades de ser inseminadas nuevamente y conseguir una preñez, comparado con vacas con un nivel de producción de leche menor al promedio del hato.

Asimismo, Jorjani (2005) observó que el intervalo entre la primera y la última inseminación es una medida adecuada de la tasa de concepción en vacas lecheras, al no estar compuesto por el intervalo parto-primer servicio, período que está afectado principalmente por decisiones del productor. Esto se debe a que una alta tasa de concepción se relaciona con un más corto intervalo primer-último servicio. Sin embargo, ocurre generalmente que un alto porcentaje de vacas logran la preñez en el último servicio, dando lugar a una distribución sesgada para el carácter. Por tanto, Schneider et al. (2006) analizaron dicho indicador reproductivo mediante simulación, a los efectos de evaluar si el ajuste del intervalo primer-último servicio empleando modelos de supervivencia permitía obtener estimaciones más precisas de la tasa de concepción, en comparación análisis basados en un modelo lineal. De este modo, los autores reportaron una correlación de 0,803 entre los verdaderos valores de cría para tasa de concepción y aquellos predichos para intervalo primer-último servicio ajustando un modelo de supervivencia, en comparación con la correlación de 0,744 obtenida mediante el ajuste de un modelo lineal. Este resultado sugiere que, cuando se tiene información del estado de preñez, los modelos de supervivencia serían más precisos que los modelos lineales al momento de predecir valores de cría para la tasa de concepción, teniendo en cuenta registros de intervalo primer-último servicio.

De forma más general y retomando el trabajo de Mark (2004), es importante conocer y divulgar el estado de los sistemas nacionales de evaluaciones genéticas, con el fin de disponer de un escenario adecuado para la aplicación de dichas pruebas a nivel internacional. A tal efecto, el autor discutió los resultados publicados en diferentes países relacionados con caracteres funcionales (Groen et al., 1997), encontrando que el intervalo en días desde la primera a la última inseminación fue empleado en diferentes países como indicador de la habilidad de la vaca para lograr una preñez.

### **2.1.1.2. Días Abiertos**

Conceptualmente, los días abiertos se definen como el período transcurrido entre el parto y el inicio de una nueva preñez. Como se mencionó, este carácter está determinado principalmente por el periodo voluntario de espera, la tasa de detección de celos y la tasa de concepción. En el caso del periodo de espera voluntario, que ha sido uno de los primeros indicadores del desempeño reproductivo de vacas lecheras, está relacionado con el tiempo de involución uterina y con el inicio de la ciclicidad ovárica. Al respecto, Whitmore et al. (1974) evaluaron si un servicio postparto temprano resultaba perjudicial en el desempeño reproductivo futuro de la vaca. Los autores observaron que vacas servidas en el primer estro postparto tuvieron menos días abiertos que hembras servidas en celos posteriores.

Estadísticamente, el carácter días abiertos fue analizado recurriendo a diferentes metodologías, desde modelos de regresión polinomial (Gill and Allaire, 1976), mínimos cuadrados (Silva et al., 1992), u otros análisis más complejos como el de supervivencia (Lee et al., 1989). Estos últimos autores evaluaron la conveniencia del análisis de supervivencia como un método para identificar diversos factores que afectan el desempeño reproductivo en los hatos. De acuerdo con los resultados, el análisis de supervivencia fue una herramienta acorde para cuantificar y evaluar los días abiertos, principalmente por la factibilidad de incluir observaciones censuradas.

Para estimar los parámetros genéticos de dispersión, Silva et al. (1992) evaluaron los factores genéticos y ambientales que afectan el desempeño reproductivo de vacas Holstein, Jersey y Guernsey. Los citados autores observaron efectos negativos de la variable estación para días abiertos, dando lugar a incrementos en dicho carácter superiores a 12 días. Además, estimaron valores de heredabilidad y repetibilidad de 0,05 y 0,06 respectivamente, y correlaciones genotípicas iguales a 0,85 y 0,96 con el intervalo parto-primer servicio e intervalo entre partos, respectivamente. No obstante, los autores recomendaron investigaciones adicionales para determinar si la selección por dichos caracteres tenía justificación económica, aspecto que será desarrollado en la sub-sección siguiente.

### **2.1.1.3. Importancia económica de los caracteres reproductivos**

Louca y Legates (1968) evaluaron el impacto de los días abiertos en la productividad del hato, indicando que vacas de primer parto debían tener un periodo abierto cercano a los 120 días para maximizar la producción. Asimismo, Schaeffer y Henderson (1972) aportaron datos donde la producción por día se incrementó a medida que los días abiertos fueron cercanos a 120 días. Con lo cual, el porcentaje de días abiertos resulto relevante en la variabilidad del desempeño total durante la vida productiva y por tanto, en la productividad del hato.

Adicionalmente, Gill and Allaire (1976) estudiaron la relación entre una función de rentabilidad y caracteres de desempeño-producción (caracteres funcionales), teniendo en cuenta la vida productiva así como caracteres reproductivos, entre ellos los días abiertos. Los autores observaron que la importancia económica de los días abiertos se debió a su relación con el comportamiento reproductivo regular o cíclico necesario para obtener una lactancia. Por lo tanto, cambios en la vida productiva de la vaca lechera pueden estar afectados en gran medida por el porcentaje de días abiertos. Consecuentemente, los autores concluyeron que es escasamente conveniente mantener vacas con altas proporciones de días abiertos, debido a que conduciría a una disminución en la duración de las lactancias de relativamente alta producción.

Posteriormente, Hodel et al. (1995) sugirieron la existencia de una relación curvilínea entre caracteres de fertilidad y días abiertos, encontrando que una duración de dicho carácter no mayor a 90 días genera un incremento lineal en la fertilidad. Sin embargo, el máximo biológico estuvo cerca a los 120 días después del parto, lo cual puede estar explicado por la alta proporción de vacas que presentan celos no detectables en los periodos más extensos de días abiertos (Boichard, 1990). Al respecto, este último autor sugirió también la consideración de los costos incurridos por inseminaciones adicionales y

tratamientos veterinarios u hormonales, entre otros, dentro de las expresiones que evalúan la rentabilidad de la operación lechera.

Posteriormente, González-Recio et al. (2004) sugirieron que no existe consenso en cuanto a los caracteres de fertilidad que deben ser incluidos en los índices económicos. No obstante, los autores señalaron la conveniencia de incluir los valores económicos de caracteres de fertilidad en la construcción de índices de selección. Además, afirmaron que para evaluar la fertilidad de la vaca lechera es esencial disponer de un sistema de registros específicos para caracteres reproductivos, con el objeto de facilitar su incorporación en los programas de evaluación genética.

### **2.1.2. Vida productiva en ganado lechero.**

Otro carácter de gran importancia económica es la *vida productiva*, definida como el número de días transcurridos entre el primer parto y el descarte. Puede ser funcional (es decir, la habilidad de la vaca para permanecer en el hato sin presentar problemas sanitarios y/o reproductivos, teniendo en cuenta su nivel de producción), o verdadera, la cual hace referencia a la capacidad de la vaca para permanecer en el rodeo sin tener en cuenta las razones de posible descarte (Ducrocq, 1987; Ducrocq, et al., 1988a). Vacas con una larga vida productiva, sea funcional o verdadera, tienen menores costos de reemplazo, lo que tiende a maximizar la productividad. Sin embargo, y de acuerdo a su definición, la vida productiva funcional recibe mayor interés por parte de los mejoradores animales, dado que tiene en cuenta el efecto de la producción en el riesgo de descarte, aun teniendo baja heredabilidad (entre 0,02 y 0,08; Ducrocq, 1987).

Uno de los primeros trabajos que buscó evaluar toros lecheros a partir de las medidas de vida productiva de la progenie usando técnicas de análisis de supervivencia fue el realizado por Smith (1983), investigación posteriormente generalizada por Smith y Quaas (1984). En este último trabajo, los autores evaluaron vida productiva incluyendo los efectos genéticos, y encontraron heredabilidades para el carácter entre 0,13 a 0,16, y correlaciones genéticas positivas con producción de leche en las primeras lactancias.

De forma más general, Ducrocq et al. (1988b) sugirieron metodologías para evaluar la vida productiva en ganado lechero, basadas en el modelo semiparamétrico propuesto por Cox (1972). Además, los autores introdujeron los conceptos de vida productiva funcional y verdadera, cuya relación fue analizada por Harris et al. (1992). Los investigadores estimaron heredabilidades entre 0,02 y 0,07 para vida productiva, y correlaciones genéticas entre dicho carácter y producción de leche entre 0,37 a 0,81. Además, concluyeron que la vida productiva funcional debía ser empleada en los programas de selección en lugar de la vida productiva verdadera, debido a que el impacto de los efectos del descarte voluntario en la producción de leche es minimizado cuando se trabaja con la vida productiva funcional. Otro trabajo que integró los conceptos de vida productiva funcional y verdadera fue el realizado por VanRaden y Klaaskate (1993), quienes combinaron en su análisis registros completos y proyectados de vida productiva de vacas Holstein. Los autores estimaron heredabilidades entre 0,03 a 0,08, así como correlaciones genéticas entre vida productiva completa y proyectada entre 0,92 a 1. Asimismo, Caraviello et al. (2004b) aplicaron el análisis de supervivencia en el cálculo de la habilidad predicha de transmisión para vida

productiva en toros Jersey, mostrando la superioridad de la metodología del análisis de supervivencia con respecto al modelo lineal para el análisis de vida productiva.

En el caso de los factores relacionados con la vida productiva, Groen et al. (1997) compararon modelos con covariables tiempo-independientes versus modelos con covariables tiempo-dependientes, con el fin de determinar el efecto de la mastitis en la fecha de descarte en ganado lechero. A partir de este análisis, se concluyó que la inclusión de covariables tiempo-dependientes en el análisis permite determinar el efecto de factores de interés en eventos como el descarte de hembras lecheras.

#### **2.1.2.1. Importancia económica**

De acuerdo con Renkema y Stelwagen (1979), la rentabilidad del hato aumentó a medida que la vida productiva fue prolongada. Específicamente, los autores utilizaron un modelo económico que comparó el ingreso marginal de una vaca y su potencial reemplazo, concluyendo que el ingreso por año podía incrementarse en 20% al prolongar la vida productiva de 3,3 a 5,3 lactancias.

Al respecto, Essl (1998) observó que el valor económico de la vida productiva se debe a la disminución en el costo de reemplazo de hembras por día y al aumento en la proporción de vacas en lactancias con mayores niveles de producción. Sin embargo, Van Arendonk (1991) encontró mediante simulación estocástica, que los costos de oportunidad relacionados con el reemplazo de hembras deben ser incluidos en la estimación de las ponderaciones económicas para vida productiva para no sobreestimar la importancia económica del carácter. Adicionalmente, Golden et al. (2000) sugirieron un marco teórico como guía en la identificación de caracteres para los cuales se deben generar diferencias esperadas entre progenies en la siguiente generación de las evaluaciones lecheras nacionales. De acuerdo con estos autores, es necesario identificar caracteres indicadores que aporten información a la predicción de caracteres económicamente relevantes. Teniendo en cuenta dichos aspectos, uno de los caracteres de mayor incidencia relacionado con la fertilidad de las hembras fue la vida productiva (Snelling et al., 1995). De acuerdo con los autores, la vida productiva permite evaluar la habilidad de la vaca para continuar en producción en determinado hato, después de haber presentado un parto aproximadamente a los dos años de edad. Sin embargo, los autores resaltaron la complejidad en la medición de la vida productiva, debido a la presencia de registros pertenecientes a hembras que continúan en producción al momento del análisis. No obstante, y recurriendo a métodos como el análisis de supervivencia, es posible realizar estimaciones de vida productiva aun en presencia de observaciones incompletas, o como son definidos, datos censurados.

#### **2.1.3. Sistemas de producción de leche en Colombia.**

De acuerdo con Dirven (2001), las ganaderías de leche en América Latina están basadas en sistemas de producción extensivos, caracterizados por bajos niveles de producción en relación con los altos costos de inversión, y además, con fluctuaciones estacionales en la producción de leche. A pesar de ello, la ganadería colombiana se ha caracterizado por producir grandes avances en los últimos años, los cuales se ven reflejados en las tasas de expansión de los sistemas productivos y además, en la importancia de la

actividad lechera dentro de la economía del país. Así, en 2011 la producción láctea en Colombia se ubicó en 6452 millones de litros de leche fluida, con un incremento del 56,14% en los últimos 10 años (Ministerio de Agricultura - Fedegán, 2012).

En cuanto a los sistemas productivos, Colombia posee tanto sistemas especializados en lechería como de doble propósito (leche y carne). Holmann et al. (2003) analizaron la evolución de dichos sistemas durante el período 1991-2003 y observaron que la productividad medida como leche producida por hectárea aumentó 44% en los sistemas de doble propósito y 14% en las lecherías especializadas. Este aumento, que permitió una reducción de 16% en el costo nacional de producción de leche, puede estar explicado, entre otros factores, por el inventario ganadero en cada uno de los sistemas de producción. La población bovina total colombiana se estima en 23.048.045 millones de cabezas (Fedegan, 2012), de las cuales aproximadamente el 6% están ubicadas en ganaderías de leche. Sin embargo, de este 6%, el sistema doble propósito contabiliza el 89% del rebaño lechero, mientras que el 11% restante corresponde a sistemas especializados (Holmann et al., 2006).

La raza Holstein ha tenido gran difusión en las ganaderías de leche especializadas. Ahora bien, los cambios en las prácticas de manejo de los hatos han provocado una disminución en la vida productiva (Hare et al., 2006), además de un grave descenso de la fertilidad de la vaca (Nebel y McGilliard, 1993), por lo cual, y además de problemas por facilidad de parto y cambios en los precios de la leche que favorecen el porcentaje de sólidos totales, se ha producido un interés en el cruzamiento con otras razas (McAllister, 2002; Weigel y Barlass, 2003; VanRaden y Sanders 2003; Heins et al., 2006). Asimismo, Holmann et al. (2003) observaron que los tambos que utilizaban inseminación artificial fueron siempre más productivos que aquellos que emplearon servicio natural. Este resultado sugiere que la inseminación artificial, tanto como estrategia única de reproducción como herramienta de mejoramiento genético dentro del hato, permitió incrementar el potencial genético y por tanto, la productividad del hato.

En cuanto a características geográficas y climatológicas Colombia, al encontrarse en la zona ecuatorial y contar con el sistema montañoso de Los Andes, posee una variedad topográfica que abarca desde selvas húmedas y llanuras tropicales, hasta páramos y nieves perpetuas. Con lo cual, las variaciones climáticas no obedecen a estaciones sino a la altitud. Esta complejidad geográfica provoca grandes variaciones climáticas según la zona, ocasionando que en cada área del país las estaciones secas o lluviosas sigan diferentes ritmos. Estas condiciones serán tenidas en cuenta en el presente trabajo, al incluir en los modelos de análisis la información de hato para cada una de las hembras a evaluar.

## **2.2. Metodología del Análisis de supervivencia.**

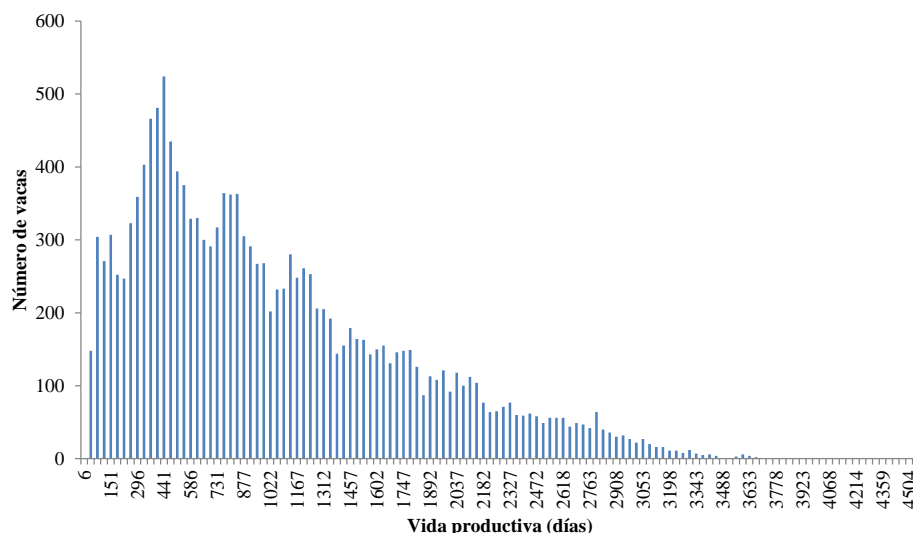
El análisis de supervivencia o análisis de datos de tiempo a la ocurrencia de un evento tiene su base en diferentes campos de estudio como la medicina, biología, ingeniería, salud pública, epidemiología y economía, entre otras. En esta metodología, la variable tiempo está enfocada en un grupo o grupo de individuos para cada uno de los cuales existe un evento definido o falla, que ocurre después de cierto periodo llamado *tiempo de falla* o *tiempo de supervivencia*. A continuación se describirán algunos antecedentes de esta metodología, específicamente en el contexto del mejoramiento genético animal.



### 2.2.1. Antecedentes del análisis de supervivencia en el mejoramiento genético animal.

Las evaluaciones genéticas basadas en su mayoría en el clásico “modelo de padre” (Henderson, 1975) presentaban varios inconvenientes para evaluar vida productiva con modelos gaussianos mixtos. Principalmente, al no tener en cuenta los descartes, se perdía información afectando la expresión de las diferencias entre padres para caracteres reproductivos de sus hijas y por tanto, causando sesgo en la estimación de los parámetros genéticos (Robertson, 1977).

Inicialmente, el análisis de caracteres relacionados con tiempo a la ocurrencia de un evento como es el caso de la vida productiva, se basó en la utilización de variables indicadoras (“proxy variables”), cuyas distribuciones satisfacían los supuestos de los modelos usados hasta el momento. Ejemplo de dichas variables fueron la permanencia productiva, el número de lactancias o la proyección del número de lactancias. En tales casos, los valores de cría predichos son obtenidos a partir de modelos unicarácter. No obstante, una particularidad general de los caracteres funcionales es que están correlacionadas negativamente con la producción de leche (Simianer et al., 1991). Parecería razonable utilizar entonces Modelos Animales Multicarácter en el análisis (Henderson y Quaas, 1976). Sin embargo, estos modelos utilizan registros parcialmente conocidos o censurados como datos faltantes, lo cual conlleva como se mencionó previamente, a una pérdida de información y a inconsistencias en la estimación. Asimismo, los datos de vida productiva tienen, usualmente, una distribución sesgada, es decir mayor frecuencia de hembras con menor vida productiva (figura 2.2.1.1). Este aspecto puede conducir a resultados incorrectos si el análisis de dichos datos se realiza con métodos que asumen normalidad (Egger-Danner, 1993). Así, surge el interés por el análisis de supervivencia (Kalbfleisch y Prentice, 1980; Cox y Oakes, 1984), definido como la metodología estadística que permite combinar la información de observaciones no censuradas (vacas que finalizaron su vida productiva) y censuradas (vacas todavía en producción al momento de la evaluación genética), considerando covariables que pueden ser o no función del tiempo y que afectan el carácter de interés.



**Figura 2.2.1.1.** Distribución de frecuencias para vida productiva (días) en Holstein Colombiano.

Dentro de la evaluación genética animal, y como se mencionó previamente, Smith y Quaas (1984) emplearon el análisis de supervivencia para predecir el valor de cría de toros lecheros teniendo en cuenta mediciones de la vida productiva de sus hijas. Posteriormente, Ducrocq (1994) y Ducrocq y Sölkner (1994) desarrollaron un programa de cálculo para ajustar modelos mediante el análisis de supervivencia. Asimismo, Schneider et al. (2002) observaron que en las evaluaciones nacionales de Francia (Ducrocq, 1999b), Austria (Ducrocq y Sölkner, 1998), Alemania (Pasman y Reinhardt, 1999) y Holanda (Harbers, 1999), se ha utilizado el análisis de supervivencia para predecir el valor de cría por vida productiva de toros Holstein y de otras razas de menor difusión. Alternativamente, Damgaard y Korsgaard, (2006a) propusieron un modelo bivariado para un carácter gaussiano y uno de supervivencia, y también para un carácter umbral y uno de supervivencia, contemplando registros faltantes (Damgaard y Korsgaard, 2006b).

Una de las familias de modelos, comúnmente utilizada en el análisis de supervivencia, es la de riesgos proporcionales (Cox, 1972). Básicamente, los modelos de riesgo proporcional (MRP) se caracterizan por tener en cuenta covariables que afectan la función de riesgo multiplicativamente (Cox, 1972). Por esto, los MRP son también conocidos como modelos de riesgo multiplicativo. Están compuestos por una función de riesgo base que depende del tiempo, y por un término positivo, el cual es una función exponencial de un vector de covariables inherentes a cada individuo, multiplicado por un vector de parámetros de regresión. La definición de los MRP implica, según Jenkins (2004), que diferencias absolutas en el vector de covariables generan diferencias proporcionales en el riesgo en determinado tiempo. Asimismo, Prentice y Kalbfleisch (1979) remarcaron que los MRP admiten una estimación viable del parámetro de riesgo sin restringir la forma de la función de riesgo base, es decir, empleando una función arbitraria.

Otra variante que han tenidos los modelos de supervivencia, ya sean paramétricos o no, ha sido la inclusión tanto de covariables tiempo-dependientes como de efectos aleatorios, por ejemplo, el valor de cría de los individuos a evaluar. En el primer caso, la incorporación de covariables tiempo-dependientes permite tener en cuenta el cambio, a través del tiempo, de los factores que afectan el riesgo de un individuo, por ejemplo el número de partos, producción de leche, entre otros. Por su parte, extender el modelo para tener en cuenta efectos aleatorios, similar a lo que ocurre en los modelos lineales mixtos, da lugar a los modelos mixtos de supervivencia o *modelos frágiles*, siendo el efecto aleatorio definido como término *frágil* (Vaupel et al., 1979). Con respecto a esta clasificación de modelos, se han usado varios enfoques para estimar los parámetros relacionados con el término frágil, desde la implementación del algoritmo EM desarrollado por Dempster et al. (1977) (Klein, 1992; Klein et al., 1992), hasta el empleo de técnicas de Monte Carlo (Korsgaard, 1996; Korsgaard et al., 1998).

Los modelos frágiles han sido adoptados por la mayoría de países para el análisis de los caracteres de supervivencia en el ganado lechero, debido a la factibilidad de incluir registros censurados (Jamrozik et al., 2008). Al respecto, Schneider et al. (2005) indicaron por medio de simulación, que el análisis de supervivencia fue más apropiado que un modelo lineal para la predicción del mérito genético de toros, por ejemplo para el carácter tasa de concepción, teniendo en cuenta medidas de desempeño reproductivo relacionadas con intervalos de tiempo (intervalo entre el parto y el último servicio, entre otras).

## 2.2.2. Definiciones generales.

### 2.2.2.1. Censura.

Previamente se ha mencionado como una de las ventajas de ajustar los modelos de supervivencia a la inclusión de observaciones censuradas. Esta característica, se relaciona con la existencia de individuos para los cuales el tiempo a la ocurrencia del evento de interés o falla, no es observado de manera completa. Dicha observación incompleta del tiempo de falla es llamada observación censurada (Cox, 1972; Ducrocq, 1987).

En el caso del mejoramiento genético animal, una observación censurada corresponderá, entre otras situaciones, a individuos que continúan en producción al momento del análisis. Para tales individuos, la única información disponible será que el individuo ha sobrevivido hasta cierto tiempo. Como se verá posteriormente este tipo de observaciones, conocidas como censura por derecha, componen la clase más frecuente de censurado. Sin embargo, existen otros tipos de información parcial, que son objeto de estudio en el análisis de supervivencia.

a. *Censura por derecha*: Se tiene cuando la falla o evento no es observado, y por tanto, el tiempo de supervivencia es desconocido. Es decir, para el  $i$ -ésimo individuo bajo estudio, se asume que existe un tiempo de vida  $T_i$  y un tiempo de censura  $C_{id}$ , donde el subíndice  $d$  indica la censura por derecha. El tiempo de supervivencia  $T_i$  será conocido exactamente si y solo si,  $T_i$  es menor o igual a  $C_{id}$ . Si  $T_i$  es mayor que  $C_{id}$ , el tiempo a la ocurrencia del evento será censurado en  $C_{id}$ . Así, los datos estarán representados por pares de variables aleatorias  $(X_i, \delta_i)$ , donde  $\delta_i$  indica si la variable  $T_i$  corresponde a un evento observado ( $\delta_i = 1$ ) o censurado ( $\delta_i = 0$ ). Además,  $X_i$  será igual a  $T_i$  si el tiempo de falla es observado, o igual a  $C_{id}$  si es censurado, es decir,  $X_i = \min(T_i, C_{id})$  (Klein y Moeschberger, 2003).

De acuerdo con Ducrocq (1987), existen ciertas situaciones, como fin del estudio o pérdida de seguimiento, que pueden conducir a observaciones censuradas por derecha:

- Censura Tipo I: El evento se observa solo si ocurre antes de un punto predeterminado. Todos los tiempos de censura  $C_i$  serán iguales a  $k$ , donde  $k$  es una constante (tiempo de finalización). Es decir, todas las observaciones censuradas tendrán tiempos iguales a la longitud del período de estudio (Cox, 1972).
- Censura Tipo II: Ocurre cuando el seguimiento es detenido después de observar un determinado número de fallas  $d$  (Cox, 1972). Es decir, el estudio continúa hasta la falla de los primeros  $d$  individuos, donde  $d$  es un entero predefinido ( $d < n$ ).
- Riesgos en competencia: Ocurre cuando individuos bajo estudio experimentan un evento diferente al de interés, causando su retiro del estudio. Es decir, el evento de interés no es observado para aquellos que experimentan el evento y serán censurados por derecha en ese punto del tiempo.

Otros tipos de censura son:

b. *Censura por izquierda*: En este caso, la falla o evento de interés ha ocurrido para el individuo antes de un tiempo determinado. Es decir, se sabe que el  $i$ -ésimo individuo ha

experimentado el evento en algún punto antes de  $C_{il}$ , (el subíndice  $l$  indica censura por izquierda), pero el tiempo exacto a la ocurrencia del evento es desconocido. El tiempo a la falla será conocido si y solo si  $T_i$  es mayor o igual que  $C_{il}$ . Las observaciones censuradas por izquierda también pueden representarse mediante pares de variables aleatorias  $(X_i, \delta_i)$ , siendo  $X_i$  igual a  $T_i$  si el tiempo a la falla es observado, por lo que  $\delta_i$  indicará si el tiempo a la falla es observado ( $\delta_i = 1$ ) o no ( $\delta_i = 0$ ). Sin embargo, y a diferencia del censurado por derecha,  $X_i = \max(T_i, C_{il})$  (Klein y Moeschberger, 2003).

c. *Doble censurado*: Se presenta cuando existe una combinación de censurado por derecha ( $C_{id}$ ) y por izquierda ( $C_{il}$ ) en el  $i$ -ésimo individuo (Turnbull, 1976). Este tipo de censura también puede representarse mediante un par de variables aleatorias  $(X_i, \delta_i)$ , donde  $X_i = \max[\min(T_i, C_{id}), C_{il}]$ . Sin embargo,  $\delta_i$  es igual a 1 si  $X_i$  es un tiempo a la muerte del individuo, igual a 0 si  $X_i$  es un tiempo de censura por derecha, e igual a  $-1$  si  $X_i$  es un tiempo de censura por izquierda. En este caso,  $T_i$  será conocido si es menor o igual a  $C_{il}$  y mayor o igual a  $C_{id}$  (Klein y Moeschberger, 2003).

d. *Censura en el intervalo*: Un tipo más general de censura ocurre cuando el tiempo a la falla es solo conocido por ocurrir dentro de un intervalo. Es decir, para el  $i$ -ésimo individuo, se sabe que el tiempo a la ocurrencia del evento de interés ocurrió dentro del intervalo  $[L_i, R_i]$ , siendo  $L_i$  y  $R_i$  el límite izquierdo y derecho del intervalo respectivamente.

e. *Truncamiento*: ocurre cuando solo se observan individuos cuyo tiempo a la falla está dentro de cierto intervalo ( $Y_L, Y_R$ ). Individuos cuyo tiempo a la ocurrencia del evento no está en dicho intervalo, no serán observados y por tanto, no tendrán información disponible. Esta es la diferencia con respecto al censurado, ya que en este último, se tendrá ya sea información completa (no censurados) o parcial (censurados) de todos los individuos.

### 2.2.2.2. Descripción de los datos de supervivencia.

Otro aspecto importante en los modelos de supervivencia es la especificación del tiempo de supervivencia o de falla. A tal efecto, se debe tener en cuenta tanto la escala de medición, como una definición clara del punto de origen y del significado de la variable tiempo de interés. De acuerdo con Cox (1972), el punto de origen debe ser definido de manera exacta para cada individuo, resultando conveniente que el tiempo de inicio de todos los individuos sea tan comparable como sea posible. Este punto puede no ser necesariamente el momento en el cual un individuo entra al estudio. Por otra parte, y aunque en general la escala de medición del tiempo es el tiempo real, también puede recurrirse a tiempo operativo (días julianos, por ejemplo). Una de las razones para la elección de la escala de tiempo es que permite el estudio de individuos en características similares. Es decir, dos individuos manejados de forma similar deberían estar en un punto equivalente después de un lapso igual de tiempo. Finalmente, es importante definir claramente el tiempo a la falla, el cual dependerá del contexto del estudio y estará relacionado con la ocurrencia de un evento específico (muerte, descarte, incidencia de una enfermedad, entre otros). En algunas áreas, no existe arbitrariedad en la definición del evento de interés. Sin embargo, del significado del tiempo a la ocurrencia del evento dependerá si el estudio se centra en el tiempo a la falla o en el análisis del desempeño total medido como una función del tiempo.

### 2.2.3. Distribuciones de los tiempos de supervivencia.

Conceptualmente, sea  $T$  una variable aleatoria no negativa que representa el tiempo a la ocurrencia del evento de interés o tiempo de falla (Kleinbaum, 1996). En el caso de vida productiva, dicho evento hace referencia al descarte de una vaca. Dicha variable  $T$ , con valor realizado  $t$ , puede describirse mediante las siguientes funciones (Smith, 2002; Lawless, 1982):

*Función de supervivencia:* Definida como  $S(t) = P(T \geq t) = 1 - F(t)$  para todos los valores observados de la variable aleatoria  $T$ , describe la fracción de individuos que continúan vivos al tiempo  $t$ . Está relacionada con la Función de distribución acumulada de  $T$ ,  $F(t)$ .

*Función de densidad:* Relacionada con el riesgo instantáneo de falla en el tiempo  $t$ , es:

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T \leq t + dt)}{dt} = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad [1]$$

Además y de acuerdo con [1],  $f(t)$  está relacionada con  $F(t)$  y  $S(t)$ .

*Función de riesgo:* Para la variable aleatoria de supervivencia  $T$ , la función de riesgo  $h(t)$  definida para  $t > 0$ , está dada por:

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T \leq t + dt | T \geq t)}{dt} = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt} \quad [2]$$

La expresión  $h(t)$  representa la tasa instantánea de falla o muerte en el tiempo  $t$ , condicional en la supervivencia hasta el tiempo  $t$ . Contrastando con  $S(t)$  – medida centrada en la no ocurrencia del evento –,  $h(t)$  focaliza en la falla. Lawless (1982) observó que esta función tiene un rol importante en el análisis de supervivencia dado que, en muchos casos, la naturaleza exacta de  $f(t)$  o  $S(t)$  es desconocida. Sin embargo, es frecuente contar con información del modo en que  $h(t)$  cambia en el tiempo. Por lo tanto, y recordando que  $f(t)$ ,  $S(t)$  y  $h(t)$  están vinculadas, el conocimiento de  $h(t)$  definirá implícitamente las demás funciones. Por ejemplo, a partir de la expresión para  $h(t)$  en [2], es posible obtener la función de supervivencia:

$$S(t) = \exp\left(-\int_0^t h(u) du\right) \quad [3]$$

*Función de riesgo acumulado:* es denotada por  $H(t)$  y mide el riesgo total (es decir, el integrado) desde el momento 0 hasta el tiempo  $t$ :

$$H(t) = \int_0^t h(u) du \quad \Rightarrow S(t) = \exp[-H(t)] \quad [4]$$

Como se verá posteriormente, tanto la función de riesgo  $h(t)$  como la de supervivencia  $S(t)$ , son relevantes por ser la base de la función de verosimilitud.

## 2.2.4. Métodos de estimación.

### 2.2.4.1. Estimación bajo Modelos paramétricos

En análisis de supervivencia, existen varios enfoques que permiten estimar los parámetros de interés. Uno de ellos es la utilización de métodos paramétricos, que involucran la maximización de la función de verosimilitud relacionada con la información disponible (Smith, 2002). En esta situación se asume un modelo probabilístico cuya forma dependerá, entre otros factores, de la conveniencia de la inferencia estadística, así como del comportamiento de la función de riesgo y de supervivencia cuando son ajustados a la variable de estudio (Cox, 1972).

De acuerdo con Lawless (1982), los tiempos de censura son frecuentemente aleatorios (tal como se describe en el apartado 2.2.9.1). Esto, debido a que el supuesto de censura aleatoria es útil, entre otras cosas, al momento de realizar inferencia estadística de procedimientos que involucran datos de tiempos de vida. Con lo cual, asumiendo dicho supuesto, las observaciones de tiempos a la ocurrencia de un evento estarán representadas por las variables  $Y_i$  y  $\delta_i$  (donde  $Y_i = \min(T_i, C_i)$  y  $\delta_i = \text{Ind}(T_i < C_i)$ ). En tal caso, sean  $C_1, C_2, \dots, C_n$  los tiempos de censura i.i.d. Dichos momentos pueden ser descritos mediante la función de densidad  $g(c)$  y de la función de supervivencia  $G(c) = \text{Prob}(T > t)$ . De esta forma, la contribución a la función de verosimilitud  $L_{i(j)}$  para una observación **no censurada**  $y_i$  ( $j=1$ ) será igual a:

$$\begin{aligned} L_{i(1)} &= \text{Prob}[Y_i \in (y_i, y_i + dt); \delta_i = 1] \\ &= \text{Prob}[T_i \in (y_i, y_i + dt); C_i \geq y_i] \\ &= \text{Prob}[T_i \in (y_i, y_i + dt)] \times \text{Prob}[C_i \geq y_i] \\ &= f(y_i) \times G(y_i) \end{aligned} \quad [5]$$

Y para una observación **censurada**  $y_i$  ( $j=0$ ):

$$\begin{aligned} L_{i(0)} &= \text{Prob}[Y_i \in (y_i, y_i + dt); \delta_i = 0] \\ &= \text{Prob}[C_i \in (y_i, y_i + dt); T_i \geq y_i] \\ &= \text{Prob}[C_i \in (y_i, y_i + dt)] \times \text{Prob}[T_i \geq y_i] \\ &= g(y_i) \times S(y_i) \end{aligned} \quad [6]$$

Definiendo  $\beta$  al vector de parámetros a estimar para determinada distribución, la función de verosimilitud es la siguiente:

$$L(\boldsymbol{\beta}) = \prod_{i \in \{\text{no cens}\}} f(y_i) G(y_i) \prod_{i \in \{\text{cens}\}} g(y_i) S(y_i) \quad [7]$$

Sin embargo, y debido a que la censura es aleatoria:

$$L(\boldsymbol{\beta}) \propto \prod_{i \in \{\text{no cens}\}} f(y_i) \prod_{i \in \{\text{cens}\}} S(y_i) \propto \prod_{\{\text{todos } i\}} [f(y_i)]^{\delta_i} [S(y_i)]^{1-\delta_i} \quad [8]$$

Dado que  $f(y_i) = h(y_i) S(y_i)$ , tenemos que:

$$L(\boldsymbol{\beta}) \propto \prod_i \left\{ [h(y_i)]^{\delta_i} S(y_i) \right\} \quad [9]$$

O de manera equivalente:

$$\log L(\boldsymbol{\beta}) = k + \sum_{i \in \{\text{no cens}\}} \log h(y_i) + \sum_i \log S(y_i) \quad [10]$$

Con lo cual, y de acuerdo con Lawless (1944), el vector de parámetros  $\boldsymbol{\beta}$  puede ser estimado maximizando [9] o su logaritmo [10].

#### 2.2.4.2. Estimación bajo Modelos no paramétricos.

Uno de los puntos clave de los métodos de estimación descritos previamente es el requerimiento de la especificación de una distribución para los tiempos de supervivencia. Sin embargo, existen otras alternativas de estimación que no requieren dicho supuesto.

##### 2.2.4.2.1. Estimador del producto límite o Kaplan Meier

Desarrollado por Kaplan y Meier (1958), permite estimar la función de supervivencia  $S$ , teniendo en cuenta datos censurados por derecha. A tal efecto, sean  $T_1, T_2, \dots, T_n$  los distintos tiempos de falla censurados por derecha, con tiempos de censura iguales a  $C_1, C_2, \dots, C_n$ . Las observaciones para los individuos consisten en el par  $(Y_i, \delta_i)$   $i = 1, 2, \dots, n$ , y donde  $Y_i = \min(T_i, C_i)$  y  $\delta_i = 1$  si  $T_i \leq C_i$ , o  $\delta_i = 0$  si  $T_i > C_i$ .

Asumiendo inicialmente que no hay “empates” (tiempos en los que fallan más de una observación), sean  $T_{(1)} < T_{(2)} < T_{(3)} < \dots < T_{(n)}$  los estadísticos ordenados para los distintos tiempos de falla no censurados. Para construir el estimador de supervivencia  $S$ , es conveniente dividir el soporte de la distribución de los tiempos de falla observados  $(0, T_{(n)})$ , en intervalos  $(T_{(j-1)}, T_{(j)}]$ , con  $j = 1, 2, 3, \dots, n$  tal que  $T_0 = 0$ . A partir de esto, sea  $R(T_{(j)})$  el conjunto de individuos en riesgo en el tiempo  $j$ ;  $N_j$  el número de individuos en riesgo en  $R(T_{(j)})$ ;  $D_j$  el número de fallas observadas en  $T_{(j)}$ ; y  $p_j$  la probabilidad condicional  $P(\text{sobrevivir a } T_j | \text{vivir en } T_{(j-1)})$ . Entonces, el estimador Kaplan-Meier de la función de supervivencia  $S(t)$  puede obtenerse como el producto de las probabilidades condicionales  $p_j$  como sigue:

$$S(t) = \text{Prob}[T \geq t] = \text{Prob}[T \geq T_{(1)}] \times \text{Prob}[T \geq T_{(2)} | T \geq T_{(1)}] \cdots \times \text{Prob}[T \geq T_{(k)} | T \geq T_{(k-1)}] \quad [11]$$

Una alternativa para estimar cada una de las probabilidades condicionales, teniendo en cuenta  $N_j$  y  $D_j$  es

$$\hat{p}_j = \frac{\text{Número de individuos que permanecen vivos después de } T_{(k)}}{\text{Número de individuos en riesgo de fallar en } T_{(k)}} = \frac{N_k - D_k}{N_k} \quad [12]$$

Combinando [11] y [12], el estimador no paramétrico o estimador Kaplan-Meier de  $S(t)$  es:

$$\hat{S}(t) = \hat{S}_{KM}(t) = \prod_{k|T_{(k)} \leq t} \left( \frac{N_k - D_k}{N_k} \right) \quad [13]$$

#### 2.2.4.2.2. Test de hipótesis bajo estimación no paramétrica

En ocasiones puede ser de interés evaluar si los estimadores Kaplan-Meier son estadísticamente equivalentes en diferentes poblaciones, verificando que las distintas curvas de supervivencia sean o no diferentes. Es entonces necesario un test estadístico que permita comparar las curvas de  $S(t)$  bajo el estimador Kaplan-Meier de diferentes poblaciones. Uno de los test más empleado es el log-rank test, que es una prueba chi-cuadrado para muestras grandes y que emplea un estadístico que proporciona una comparación general de las curvas para las poblaciones evaluadas. Dicho estadístico, al igual que otros test chi-cuadrado, hace uso de las diferencias entre los datos observados versus los esperados para todas las poblaciones. A tal fin, si se define como  $\tau$  al mayor tiempo de falla en el estudio, las hipótesis planteadas son las siguientes:

$$H_0: h_1(t) = h_2(t) = \cdots = h_K(t) \text{ para todo } t \leq \tau.$$

$$H_1: \text{Al menos uno de los } h_j(t) \text{ es diferente para algún } t \leq \tau.$$

En este caso, se rechazará la hipótesis nula si al menos una de las poblaciones bajo estudio es diferente de las demás en algún  $t$ . Entonces, sean los tiempos de falla ordenados de manera creciente para todas las sub-poblaciones en conjunto ( $t_1 < t_2 < t_3 \cdots < t_d = \tau$ );  $d_{ij}$  es el número de fallas observadas en el tiempo  $t_i$  en la sub-población  $j$ ; y  $n_{ij}$  el número de individuos en riesgo de fallar en la  $j$ -ésima sub-población. Asimismo, sea  $W_j(t_i)$  la función de ponderación para los diferentes tiempos, la cual permitirá dar mayor o menor peso a los diferentes tiempos de falla, y sea  $D$  el número total de tiempos de falla observados. Entonces, si la hipótesis nula es verdadera, el estimador del riesgo esperado en la  $j$ -ésima sub-población en el tiempo  $i$  es igual a  $d_i/n_i$  y el estadístico de prueba será igual a:

$$Z_j(\tau) = \sum_{i=1}^D W_j(t_i) \left\{ \frac{d_{ij}}{n_{ij}} - \frac{d_i}{n_i} \right\}, j = 1, \dots, K \quad [14]$$



En [14], la función de ponderación ( $W(t_i)$ ) es igual a 1 para todos los tiempos de supervivencia. Para rechazar  $H_0$ , al menos uno de los  $Z_j(\tau)$  debe estar alejado de cero, y por tanto, la  $j$ -ésima sub-población tendrá un riesgo diferente al esperado bajo la hipótesis nula. De lo contrario, y si los valores de  $Z_j(\tau)$  para todas las sub-poblaciones son cercanos a cero, existirá poca evidencia para rechazar  $H_0$ . Es decir, la hipótesis nula que será testeada es que no hay diferencia entre las curvas de supervivencia para las  $K$  poblaciones. Además bajo  $H_0$ , si el número de grupos comparados es  $K$  ( $\geq 2$ ), el estadístico log-rank es aproximadamente chi-cuadrado con  $K-1$  grados de libertad.

### 2.2.5. Modelos de regresión.

Dentro del análisis de supervivencia es de interés relacionar el riesgo de descarte de una vaca con ciertas variables como el número de partos, o el estado de lactancia, entre otros. Si se define  $\mathbf{x} = (x_1, \dots, x_n)'$  como el vector de variables explicativas y  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$  al vector de parámetros de regresión, la función de riesgo de una vaca con vector de variables  $\mathbf{x}$  se denominará ahora  $h(t|\mathbf{x})$ . Sin embargo, existen varias formas de asociar  $h(t)$  con el vector  $\mathbf{x}$ , dentro de las cuales se encuentran los modelos de tiempo acelerado y los modelos de riesgo proporcional. El ajuste de uno u otro modelo dependerá de los supuestos hechos para escribir  $h_0(t)$ .

#### 2.2.5.1. Modelos de tiempo de falla acelerada.

Estos modelos establecen que el vector de covariables  $\mathbf{x}$  actúa tanto multiplicativamente en el tiempo de falla  $t$ , como también generando un cambio en la escala del tiempo de la función de riesgo base  $h_0(t)$ , dado que  $t$  es ahora  $t^* = \exp(\mathbf{x}'\boldsymbol{\beta})t$ :

$$h(t|\mathbf{x}) = h_0(\exp(\mathbf{x}'\boldsymbol{\beta})t) \exp(\mathbf{x}'\boldsymbol{\beta}) = h_0(t^*) \exp(\mathbf{x}'\boldsymbol{\beta}) \quad [15]$$

Con lo cual, y de acuerdo con Ducrocq (2001), si  $\mathbf{x}'\boldsymbol{\beta} > 0$ , entonces  $\exp(\mathbf{x}'\boldsymbol{\beta}) > 1$ , dando lugar a un tiempo acelerado. Por el contrario, si  $\mathbf{x}'\boldsymbol{\beta} < 0$ , entonces  $\exp(\mathbf{x}'\boldsymbol{\beta}) < 1$ , siendo entonces un tiempo desacelerado.

#### 2.2.5.2. Modelos de riesgo proporcional.

En los Modelos de Riesgo Proporcional (MRP; Cox, 1972; Lawless, 1982), las variables afectan la función de riesgo multiplicativamente. Esto implica que la función de riesgo de  $T$  dado  $\mathbf{x}$ , puede ser escrita como:

$$h(t|\mathbf{x}) = h_0(t) g(\mathbf{x}) \quad [16]$$

En [16], tanto  $h_0$  como  $g$  pueden involucrar parámetros desconocidos, siendo  $h_0(t)$  la función de riesgo base para un individuo con  $g(\mathbf{x})=1$ . De esta forma en los MRP, los diferentes individuos tendrán funciones de riesgo que son proporcionales entre si. Es decir, el cociente  $h(t|\mathbf{x}_1) / h(t|\mathbf{x}_2)$  de las funciones de riesgo para dos individuos con vectores de regresión  $\mathbf{x}_1$  y  $\mathbf{x}_2$  no varía con el tiempo. Este supuesto de proporcionalidad permite, entre otros aspectos, interpretar el efecto multiplicativo de las covariables en la función de riesgo de los individuos bajo estudio (Cox, 1972).

Uno de los MRP más frecuentemente usado es el modelo Weibull. Asumiendo dicha distribución,  $h(t|\mathbf{x})$  se define como el producto de una función de riesgo base  $h_0(t)$ , la cual describe el proceso natural de envejecimiento de la población, multiplicada por una función exponencial paramétrica  $\exp[\mathbf{x}'\boldsymbol{\beta}]$ . Es decir:

$$h(t|\mathbf{x}) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}) \quad [17]$$

Por otro lado, si el riesgo es constante en el tiempo, se define un modelo de riesgo exponencial, donde  $h_0(t) = \lambda$ . Por último, y si  $h_0(t)$  es completamente arbitraria, el modelo de riesgo proporcional es conocido como un “Modelo de Cox” (Cox, 1972) o modelo semi-paramétrico. En este último caso, se dice que es semi-paramétrico porque si bien  $h_0(t)$  no sigue una distribución particular, el término exponencial es función de un vector de parámetros. Así, y retomando el enfoque paramétrico Weibull para la función de riesgo base  $h_0(t)$ , la expresión [17] es ahora:

$$h(t|\mathbf{x}) = \lambda \rho (\lambda t)^{\rho-1} \exp[\mathbf{x}'\boldsymbol{\beta}] \quad [18]$$

donde,  $\lambda \rho (\lambda t)^{\rho-1}$  es la función de riesgo base Weibull, con parámetros de forma y escala,  $\lambda$  y  $\rho$ , respectivamente.

### 2.2.5.3. Estimación paramétrica en modelos de regresión.

Siguiendo el enfoque de Lawless (1982) y teniendo en cuenta un modelo de regresión Weibull, el logaritmo de la verosimilitud para modelos paramétricos, definiendo  $y_i = \min(T_i, C_i)$  es:

$$\log \ell(\boldsymbol{\beta}) = \text{constante} + \sum_{i \in \{\text{no cens}\}} \log h(y_i) + \sum_i \log S(y_i) \quad [19]$$

donde, para un modelo de regresión Weibull se tiene que:

$$\log S(y_i) = -y_i^\rho \exp\{\mathbf{x}_i'\boldsymbol{\beta}\} \quad [20]$$

y

$$\log h(y_i) = \log(\rho y_i^{\rho-1} \exp(\mathbf{x}_i'\boldsymbol{\beta})) \quad [21]$$

Si omitimos el término constante, el logaritmo de la función de verosimilitud resultante es:

$$\log \ell(\rho, \boldsymbol{\beta}) = N \log \rho + (\rho - 1) \sum_{i \in \{\text{no cens}\}} \log y_i + \sum_{i \in \{\text{cens}\}} \mathbf{x}_i'\boldsymbol{\beta} - \sum_i y_i^\rho \exp(\mathbf{x}_i'\boldsymbol{\beta}) \quad [22]$$

En este caso,  $N$  es el número total de observaciones no censuradas. A partir de estas expresiones, es posible encontrar el estimador de máxima verosimilitud de  $(\rho, \boldsymbol{\beta})$ .

#### 2.2.5.4. Estimación semi-paramétrica: modelo de Cox.

En los modelos de Cox, la función de riesgo de los individuos consiste en el producto de una función de riesgo base  $h_0(t)$  arbitraria (no paramétrica), por una función paramétrica  $\exp(\mathbf{x}'\boldsymbol{\beta})$ , dando lugar a un modelo semi-paramétrico. Para estimar  $\boldsymbol{\beta}$  es conveniente emplear un enfoque similar al utilizado para el estimador Kaplan-Meier. Por tanto, sean  $T_{[1]} < T_{[1]} < \dots < T_{[m]}$  los distintos tiempos de falla no censurados ordenados. Se asume que en cada tiempo  $T_{[k]}$  existe solo una falla, y además que  $R(t)$  es el conjunto de individuos en riesgo al  $t$ . Así, la probabilidad condicional de falla para el  $i$ -ésimo individuo en el tiempo  $T_{[k]}$ , dada por el conjunto en riesgo en  $T_{[k]}$  y la ocurrencia de la falla en dicho tiempo, es igual a:

$$\begin{aligned} \text{Prob}\left[i \text{ falle en } T_{[k]} \mid R(T_{[k]}), \text{ una falla ocurre en } T_{[k]}\right] &= \frac{\text{Prob}\left[i \text{ falle en } T_{[k]}\right]}{\sum_{j \in R(T_{[k]})} \text{Prob}\left[j \text{ falle en } T_{[k]}\right]} \\ &= \frac{\exp\{x_i'\boldsymbol{\beta}\}}{\sum_{j \in R(T_{[k]})} \exp\{x_j'\boldsymbol{\beta}\}} \end{aligned} \quad [23]$$

Combinando la anterior probabilidad condicional sobre todas las fallas no censuradas, y definiendo al individuo  $i$  que falla en  $T_{[k]}$  como  $k$ , la verosimilitud parcial de Cox es:

$$\ell(\boldsymbol{\beta}) = \prod_{k \in \{\text{no cens}\}} \left[ \frac{\exp\{x_k'\boldsymbol{\beta}\}}{\sum_{j \in R(T_{[k]})} \exp\{x_j'\boldsymbol{\beta}\}} \right] \quad [24]$$

Entonces se puede estimar  $\boldsymbol{\beta}$  maximizando la expresión [24].

#### 2.2.6. Generalizaciones.

##### 2.2.6.1. Covariables tiempo - dependientes.

En los modelos de supervivencia ciertos efectos fijos que afectan el desempeño de un individuo son independientes del tiempo. Tal es el caso del sexo, la raza y el hato, entre otros. Sin embargo, existen otros efectos fijos que afectan diferencialmente la variable respuesta según el tiempo de evaluación. Considérese, por ejemplo, el número de partos. Es razonable suponer que cuantos más partos haya tenido una vaca hasta un determinado momento, mayor será la probabilidad de que sea descartada. Así, la misma vaca tendrá más posibilidad de ser rechazada si ya alcanzó su tercer parto, por ejemplo, que cuando no lo presentó. Este tipo de efecto fijo recibe el nombre de efecto fijo tiempo-dependiente. A modo de ejemplo, emplearemos el efecto número de partos. Para dicho factor, puede definirse un efecto tiempo-dependiente indicador  $x_{ik}(t)$  para la  $i$ -ésima vaca y el  $k$ -ésimo número de parto (nivel del efecto) como:

$$x_{ik}(t) = \begin{cases} 1 & \text{si al tiempo } t \text{ la vaca está entre su } (k-1) \text{ y su } k\text{-ésimo parto.} \\ 0 & \text{en caso contrario.} \end{cases} \quad [25]$$

En este contexto,  $t$  representa un punto en el tiempo que transcurre desde el primer parto hasta el descarte del animal. Si por ejemplo en este tiempo la  $i$ -ésima vaca presentó tres partos,  $\mathbf{x}'_i(t)\boldsymbol{\beta}$  modela el efecto tiempo-dependiente “número de partos” mediante una función del tiempo constante por partes (*Piecewise constant function*). En ese caso,

$$\mathbf{x}_i(t) = \begin{cases} (1 & 0 & 0) & \text{si al tiempo } t, \text{ la vaca ha presentado su } 1^{\text{er}} \text{ parto} \\ (0 & 1 & 0) & \text{si al tiempo } t, \text{ la vaca ha presentado su } 2^{\text{do}} \text{ parto} \\ (0 & 0 & 1) & \text{si al tiempo } t, \text{ la vaca ha presentado su } 3^{\text{er}} \text{ parto} \end{cases} \quad [26]$$

#### 2.2.6.2. Estratificación.

Kalbfleisch y Prentice (2002) observaron que la dependencia del tiempo por parte de las variables regresoras puede considerarse mediante la inclusión de estratos, hecho que facilita la extensión del análisis de supervivencia en situaciones donde el supuesto de riesgo proporcional no es válido para todos los efectos, o durante todo el período de tiempo analizado. De acuerdo con Aalen et al. (2008), la población es dividida en  $k$  estratos, por lo que para el individuo  $i$  en el estrato  $s$ , la función de riesgo es la siguiente:

$$h(t | \mathbf{x}_i) = h_{s0}(t) \exp(\mathbf{x}_i(t)\boldsymbol{\beta}) \quad [27]$$

En [27], los efectos de las covariables se asumen constantes a través de estratos, a diferencia del riesgo base que varía entre ellos. En cuanto a la estimación del vector de coeficientes de regresión  $\boldsymbol{\beta}$ , la expresión de la verosimilitud parcial incorporando la estratificación es:

$$L(\boldsymbol{\beta}) = \prod_{s=1}^k \prod_{T_{sj}} \frac{\exp[\mathbf{x}_{i_j}(T_{sj})\boldsymbol{\beta}]}{\sum_{l \in R_{sj}} \exp[\mathbf{x}_l(T_{sj})\boldsymbol{\beta}]} \quad [28]$$

donde  $T_{s1} < T_{s2} < \dots, T_{sj}$ , son los tiempos de supervivencia observados en el estrato  $s$ , y  $R_{sj}$  es el conjunto de individuos en riesgo en dicho estrato en el tiempo  $T_{sj}$ .

#### 2.2.7. Modelos de fragilidad

Es posible incluir en los MRP, sean paramétricos o semi-paramétricos, uno o más efectos aleatorios o frágiles, dando lugar a modelos mixtos de supervivencia o modelos de fragilidad (Vaupel et al., 1979). A tal fin, sea  $\boldsymbol{\theta}$  el vector que incluye los efectos fijos y aleatorios, en este caso, un efecto aleatorio de padre  $s$ . Recurriendo a la notación de modelos lineales mixtos, para un individuo  $m$  se tiene:

$$\mathbf{w}'_m = (\mathbf{x}'_m \quad \mathbf{z}'_m) \quad \text{y} \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{s} \end{bmatrix}$$

donde  $\boldsymbol{\beta}$  es el vector de efectos fijos. La función de riesgo  $h(t)$  para el individuo  $m$  es:

$$h_m(t | \boldsymbol{\theta}, \rho) = h_0(t) v_q \exp(\mathbf{x}'_m \boldsymbol{\beta}) \quad [29]$$

siendo  $v_q$  el término aleatorio o frágil que afecta multiplicativamente el riesgo de la vaca  $m$  hija del toro  $q$ . En este caso, la transformación  $v_q = \exp(s_q)$  permite incluir el efecto aleatorio de padre en el término exponencial, con lo cual:

$$h(t | \boldsymbol{\theta}, \rho) = h_0(t) \exp(\mathbf{w}'_m \boldsymbol{\theta}) \quad [30]$$

Para el término frágil es frecuente asumir distribuciones flexibles y de conveniencia matemática, como el caso de la distribución gamma (Ducrocq, 1987; Klein, 1992) o la distribución normal inversa (Hougaard, 1986a, b). Sin embargo, se ha mostrado que las estimaciones de  $v$ , obtenidas empleando una distribución gamma son relativamente grandes, sugiriendo que dicho término frágil sigue una distribución log-normal. Teniendo en cuenta la transformación que relaciona  $v$  y  $s$ , se puede decir que  $s$  sigue aproximadamente una distribución Normal (Ducrocq, 1987; Ducrocq et al., 1988b; Ducrocq, 1994). Además, es apropiado tener en cuenta las relaciones de parentesco entre los animales, asumiendo una distribución normal multivariada para  $\mathbf{s}$ , es decir,  $v_q \sim \text{lognormal}(0, \sigma_s^2)$  y por tanto  $s_q \sim \text{Normal Multivariada}(0, \mathbf{A} \sigma_s^2)$ , donde  $\mathbf{A}$  corresponde a la matriz de relaciones de parentesco entre los padres.

Ahora, y condicionalmente en  $\boldsymbol{\theta}$  y  $\rho$ , la contribución a la función de verosimilitud del individuo  $m$ , el cual falla ( $\delta_m = 1$ ) o es censurado ( $\delta_m = 0$ ) en el tiempo  $t_m$  es:

$$\begin{aligned} \ell_m &= \ell_m(t_m | \boldsymbol{\theta}, \rho) \\ &= \left[ h(t_m | \boldsymbol{\theta}, \rho) \right]^{\delta_m} \times S(t_m | \boldsymbol{\theta}, \rho) \end{aligned} \quad [31]$$

donde  $S(t_m | \boldsymbol{\theta}, \rho)$  es la función de supervivencia en el tiempo  $t_m$ . Si se asume un riesgo base Weibull, las funciones de supervivencia y de riesgo, importantes al momento de la estimación, son:

$$\begin{aligned} h(t_m | \boldsymbol{\theta}, \rho) &= \rho t_m^{\rho-1} \exp(w'_m \boldsymbol{\theta}) \\ S(t_m | \boldsymbol{\theta}, \rho) &= \exp\left(-\int_0^{t_m} h(t | \boldsymbol{\theta}, \rho) dt\right) = \exp\left[-t_m^\rho \exp(w'_m \boldsymbol{\theta})\right] \end{aligned} \quad [32]$$

### 2.2.8. Tiempos de falla discretos – datos agrupados

En algunos casos, los datos de tiempos a la ocurrencia de un evento son discretos, ya sea por agrupamiento de los datos continuos debido a mediciones imprecisas, o debido a que la variable respuesta está expresada en unidades discretas de medición con muy pocas clases. Un ejemplo en producción animal es el número de años de vida productiva, o el número de partos en dicho período. Sin embargo, el modelo de regresión presentado en [17] no admite la presencia de empates, es decir, numerosos valores iguales de la variable respuesta. Así, surgen los modelos de datos agrupados (Prentice y Gloeckler, 1978) que permiten obtener estimaciones tanto de la función de riesgo relativo, como de la de supervivencia y de los parámetros de regresión en presencia de empates.

En el modelo de datos agrupados, los parámetros relacionados con  $h(t)$  pueden ser estimados junto con el vector de parámetros de los efectos fijos ( $\beta$ ) y con los efectos aleatorios incluidos en el modelo. A tal efecto, se define una función de verosimilitud completa y no parcial como en el modelo de Cox. No obstante, y mediante determinada reparametrización, es posible transformar el modelo de datos agrupados en uno de regresión, que incluye una covariable tiempo-dependiente adicional. Dicha covariable, que es incluida en el modelo de datos agrupados, toma distintos valores en los tiempos 0; 1; 2; 3; ... ;  $t - 1$ , es decir, cambia en cada punto del tiempo existente entre el origen y el tiempo de falla observado, o tiempo de censura. El desarrollo completo de esta metodología se describirá en el apartado de materiales y metodología de la presente tesis.

Asimismo, la inclusión de un término frágil en el modelo de datos agrupados es similar a la metodología empleada en los modelos de riesgo proporcional. El cambio sustancial ocurrirá en las funciones  $S(t)$  y  $h(t)$ , las cuales serán contempladas al momento de construir la función de verosimilitud bajo un modelo de datos agrupados.

### 2.2.9. Otros aspectos relevantes al análisis de supervivencia.

#### 2.2.9.1. Chequeo del supuesto de censura aleatoria.

El supuesto de censura aleatoria hace referencia a que, en un estudio particular en el cual la fecha de finalización es fija, el tiempo de inicio en el estudio de todos los individuos debe ser aleatorio. Es decir, los tiempos de censura  $C_i$  son observaciones de la variable aleatoria  $C$ , y por tanto, los tiempos de supervivencia serán independientes de los de censura (Smith, 2002). Así, sean  $C_1, C_2, C_3, \dots, C_n$  los tiempos de censura por derecha para las variables  $T_1, T_2, T_3, \dots, T_n$ . Si el supuesto de censura aleatoria se cumple, el tiempo de censura  $C_i$  será independiente del tiempo de falla asociado  $T_i$ . Según Lawless (1982), la censura aleatoria implica que el conocimiento de la distribución de los diferentes  $C_i$  no aportará información acerca de los tiempos de supervivencia  $T_i$ .

Aunque el cumplimiento de este supuesto genera importantes simplificaciones en el análisis estadístico, debe resaltarse que si el supuesto de censura aleatoria no es correcto, se obtendrán resultados erróneos. Sin embargo, no se tienen test adecuados que permitan verificar este supuesto. Una alternativa es incluir todos los registros censurados y no censurados y controlar posteriormente si ocurren cambios en los resultados del análisis

(Ducrocq, 2001). No obstante, resolver preguntas como la siguiente: “un individuo censurado en el tiempo  $C_i$ , ¿es representativo de los individuos que sobreviven a dicho tiempo, y que son afectados por los mismos niveles de los efectos involucrados?”, servirá como orientación al momento de asumir o no censura aleatoria (Cox, 1972).

### 2.2.9.2. Bondad ajuste modelos paramétricos.

Independientemente del modelo de supervivencia empleado (semiparamétrico o de Cox, o modelo paramétrico), el estimador del vector de parámetros  $\hat{\beta}$  se obtiene maximizando una función de verosimilitud (parcial en el caso de un modelo de Cox). Este estimador tiene las mismas propiedades que aquellos obtenidos para funciones de verosimilitud usuales. Particularmente, es asintóticamente insesgado, y sigue una distribución normal multivariada con matriz de (co)varianzas igual a la inversa del Hessiano de la función de log-verosimilitud. Esto se traduce en que asintóticamente

$\hat{\beta} \sim N\left(\beta, \left[i(\hat{\beta})\right]^{-1}\right)$ . Teniendo en cuenta la propiedad asintótica, se pueden construir

pruebas de hipótesis para probar que  $\beta = \beta_0 = 0$ . Tal es el caso de los tests de Rao y del cociente de verosimilitud. En el primer caso, la prueba de Rao evalúa si en  $\beta = 0$ , la derivada del logaritmo de la verosimilitud es diferente de 0. Además, este test se caracteriza por no requerir el cálculo explícito de  $\hat{\beta}$ . Por su parte, mediante el test del cociente de verosimilitud se comparan los valores del logaritmo de la función de verosimilitud para un modelo completo versus aquellos obtenidos para un modelo reducido. Su desarrollo y aplicación se verá detalladamente en el apartado de la metodología estadística empleada para el análisis.

### ***Capítulo 3. Materiales y Metodología.***



### **Capítulo 3. Materiales y Metodología.**

#### **3.1. Base de datos.**

Los datos para este análisis fueron recopilados desde 1982 hasta 2007 por la Asociación Holstein de Colombia. Inicialmente, la base de datos estaba compuesta por 213.034 registros de producción de leche, pertenecientes a 62351 vacas hijas de 3490 padres. Sin embargo, la información no fue completa para todas las hembras, como se observa en la tabla 1.

**Tabla 3.1.1.** Medias y desviaciones estándar para diferentes variables productivas y reproductivas.

<b>Variable</b>	<b>No. Registros</b>	<b>Media</b>	<b>Desv. Est.</b>
Días en lactancia	198972	323,49	117,85
Prod. leche (305d) Kg.	199190	6218,41	2081,45
Días secos	145689	92,03	76,26
IEP	146441	443,63	164,85
EPP	61292	1083,35	504,88
No. Partos	61292	4,72	2,25

Para los tres caracteres analizados (vida productiva, días abiertos e intervalo primer-último servicio), la información fue editada en dos etapas. En primer lugar, se descartaron registros con identificación de padre y/o hato desconocido, así como con fechas de parto faltantes. Posteriormente, se descartaron vacas con edades al primer parto menores a 18 (548 días) y mayores a 42 meses (1278 días), con partos no consecutivos, con intervalos entre partos menores a 11 meses buscando evitar inconsistencias en las fechas de parto, o que comenzaron sus registros después del primer parto. Esto último, con el fin de excluir registros censurados por izquierda. También fueron excluidos registros de hembras que presentaron un solo parto y que fueron descartadas a la parición siguiente, dado que dichas vacas no aportan ningún tipo de información a los caracteres de interés (Roxström et al. 2003). Se explicará subsiguientemente la edición que se realizó específicamente para cada carácter.

##### **3.1.1. Vida productiva (VP).**

Para el carácter vida productiva, se descartaron registros pertenecientes a hatos con menos de 5 observaciones no censuradas. Ducrocq y Quass (1988a,b) observaron que dicha política permite limitar el sesgo en la estimación del efecto hato-año. Después de la edición para VP, los datos comprendían registros pertenecientes a 15706 vacas nacidas entre 1984 y 2004, hijas de 509 toros y pertenecientes a 178 hatos. El archivo de pedigree fue construido

de modo de incluir la mayor cantidad de generaciones posibles, dando lugar a 32859 individuos en el pedigree.

La VP fue medida como el tiempo entre el primer parto y la fecha de descarte o el fin de los datos (fecha de censura). Fueron consideradas como no censuradas aquellas vacas con fecha y motivo de descarte conocidos, así también como las vacas reportadas muertas. Por su parte, las vacas con fecha de descarte desconocida fueron consideradas censuradas. De esta forma, registros pertenecientes a vacas que permanecieron vivas al final del periodo de estudio, así como las hembras con intervalos entre partos o intervalo entre el último parto y el fin del periodo de análisis mayor a 18 meses, fueron censuradas en la fecha del último parto (51.8%). En este caso, la censura se relaciona con el no reporte de un nuevo parto. Asimismo, vacas vendidas a un nuevo hato fueron censuradas en la fecha del último parto ocurrido en el primer establecimiento productivo. La exclusión de tales vacas del análisis para el segundo rodeo fue una precaución para evitar potenciales sesgos en la estimación del efecto relacionado con hato. Adicionalmente, vacas con 8 partos o más, fueron agrupadas y censuradas en la fecha de la octava parición.

**Tabla 3.1.1.1.** Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter vida productiva.

	Censurados	No censurados
Min	23	6
Max	4540	4008
Promedio	966,1	1058,3

### 3.1.2. Intervalo primer-último servicio (IPU).

Al igual que para el carácter días abiertos, la disponibilidad de información relacionada con fechas de servicio y diagnósticos de preñez permitió incluir y analizar - en este trabajo - el intervalo en días entre el primer y el último servicio reportados para un subconjunto de los animales que componen la base de datos de vida productiva. La inclusión de dicho carácter obedece a su relación con la tasa de concepción, como se mencionó en el apartado introductorio. Con lo cual, la información disponible para el análisis del carácter intervalo primer-último servicio consistió en registros para 7151 vacas, pertenecientes a 146 hatos, incluyendo animales con información de 1 a 6 partos.

De acuerdo con Schneider et al. (2006), y con el fin de evitar inconsistencias en la base de datos, vacas con un intervalo mayor a 288 días fueron eliminadas del análisis. En cuanto a la definición de censura, vacas con diagnóstico de preñez positivo fueron consideradas no censuradas a la fecha de diagnóstico, siendo su tiempo de supervivencia igual al número de días entre el primer servicio y la fecha de diagnóstico de preñez. Por su parte, vacas que no fueron reportadas como preñadas, fueron consideradas censuradas a la fecha del último servicio informado, siendo su tiempo de censura igual al número de días entre el primer y el último servicio informados.

Asimismo, el archivo de pedigree fue construido de tal forma que se incluyera la mayor cantidad de individuos, dando lugar a 32561 animales en el archivo de genealogía.

**Tabla 3.1.2.1** Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter intervalo entre el primer y último servicio.

	Censurados	No censurados
Min	0	0
Max	288	287
Promedio	72,6	51,2

### 3.1.3. Días abiertos (DA).

Teniendo en cuenta la disponibilidad de una base de datos adicional, relacionada con fechas de servicio y diagnósticos de preñez posteriores al último parto, se incluyó en el estudio el carácter días abiertos (DA) para un subconjunto de datos derivado de la base de análisis de vida productiva. Dicho carácter fue definido como el tiempo en días entre la última fecha de parto y la última fecha de servicio registrada o la fecha de diagnóstico de preñez. Para evitar inconsistencias en la información, vacas con DA menores a 20 días o mayores a 365 días fueron eliminadas del análisis (Hou et al., 2009), dando lugar a registros de días abiertos para 6205 vacas, pertenecientes a 99 hatos y con entre 1 y 6 número de partos.

En el caso de la definición de censura, vacas diagnosticadas preñadas fueron consideradas no censuradas a la fecha del diagnóstico, siendo su tiempo de supervivencia igual al número de días entre la fecha del último parto y la fecha de diagnóstico de preñez. Por su parte, vacas que no fueron reportadas como preñadas, fueron censuradas a la fecha del último servicio informado, siendo su tiempo de censura igual al número de días entre la fecha del último parto y la fecha del último servicio registrado.

Con el fin de incluir el efecto aleatorio de animal en el análisis, el archivo de pedigree fue construido de manera tal que sea lo más completo posible, e incluyó 31952 animales.

**Tabla 3.1.3.1.** Valores máximos, mínimos y promedio en días, para observaciones censuradas y no censuradas del carácter días abiertos.

	Censurados	No censurados
Min	21	23
Max	365	365
Promedio	166,7	138,7

## 3.2. Metodología

### 3.2.1. Análisis de supervivencia.

Sean  $T_i$  y  $C_i$  variables aleatorias que respectivamente denotan el tiempo de falla u ocurrencia del evento de interés y el tiempo de censura para la vaca  $i$ ,  $i = 1, \dots, n$ . Para cada hembra se observa entonces, un conjunto de dos variables aleatorias  $(Y_i, \delta_i)$ , donde  $Y_i = \min(T_i, C_i)$  y  $\delta_i$  es una variable indicadora de censura, igual a uno si  $T_i \leq C_i$ , o cero en caso contrario. Bajo el concepto de “modelos de riesgo proporcional” (Cox, 1972), la función de riesgo de descarte de una hembra es igual al producto entre una función de riesgo base, que es común para todos los individuos incluidos en el análisis, y una función exponencial positiva de covariables de interés, las cuales pueden ser independientes o dependientes del tiempo. Es decir, la función de riesgo  $h_i(t)$  para la vaca  $i$  en el tiempo  $t$ , es igual a:

$$h_i(t) = h_0(t) \exp[\mathbf{x}_i' \boldsymbol{\beta}] \quad [33]$$

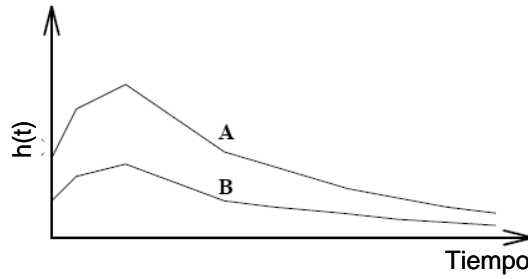
donde  $h_0(t)$  es la función de riesgo base en el tiempo  $t$ ;  $\mathbf{x}_i$  el vector de covariables tiempo-dependientes para la vaca  $i$ ; y  $\boldsymbol{\beta}$  el correspondiente vector de parámetros de regresión para los diferentes factores.

Siguiendo el enfoque planteado por Sánchez et al. (2004), la metodología de análisis de supervivencia fue empleada de manera progresiva, con el fin de controlar el supuesto relacionado con los modelos de riesgos proporcionales, el ajuste de una función de riesgo Weibull, la significancia de efectos fijos, y la inclusión y estimación de efectos aleatorios, componentes de varianza y heredabilidad. Asimismo, y siguiendo el enfoque de Calo et al. (1973) se estimaron las correlaciones genéticas entre el nivel de producción de leche y algunos de los caracteres evaluados.

#### 3.2.1.1. Supuesto de proporcionalidad

Como se mencionó en el apartado introductorio, en un modelo de riesgo proporcional el cociente de las funciones de riesgo para dos individuos, con vectores de covariables  $\mathbf{x}_1$  y  $\mathbf{x}_2$ , no variará con el tiempo pero si con las covariables que afectan a cada individuo. Por lo tanto, para verificar la necesidad de un análisis estratificado, es conveniente realizar un gráfico de  $\log(-\log(S_{KM,n}(t)))$  versus el  $\log(\text{tiempo})$  para cada uno de los estratos definidos. Dichos estratos hacen referencia a un efecto fijo tiempo independiente, como generación, edad al primer parto o rebaño, entre otros, para el cual el control del supuesto de proporcionalidad sea de interés e indique la conveniencia o no de ajustar una función de riesgo base común para todos los individuos en la población. En este caso, el término  $S_{KM,n}(t)$  hace referencia al estimador Kaplan-Meier de la función de supervivencia en el tiempo  $t$  para el  $n$ -ésimo nivel del efecto definido como estrato.

Así, para un determinado conjunto de datos y si el supuesto de proporcionalidad se cumple, el resultado de graficar el  $\log(-\log(S_{KM,n}(t)))$  versus el  $\log(\text{tiempo})$  para cada uno de los estratos, debe mostrar paralelismo entre los ajustes. A modo de ejemplo, se ilustra en la figura 3.2.1.1.1 el supuesto de proporcionalidad para dos individuos A y B, caso en el cual las funciones de riesgo de dichos individuos son proporcionales (Ducrocq, 2001).



**Figura 3.2.1.1.1.** Ilustración del supuesto de riesgos proporcionales.

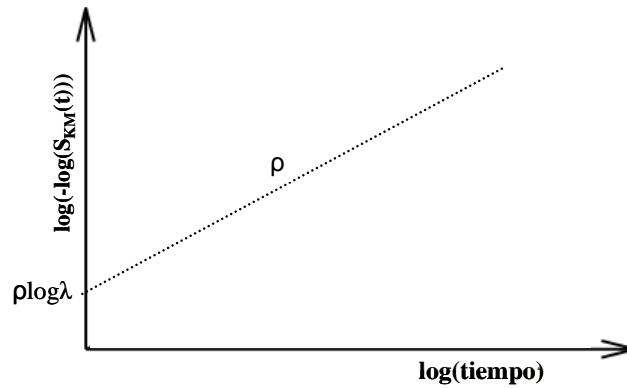
### 3.2.1.2. Distribución Weibull para la función de riesgo base

Asumir una forma paramétrica para la función de riesgo base permite, en ciertos casos, modelar el comportamiento de una variable de tiempo de manera menos compleja, así como también simplificar la matemática del análisis. Como se mencionó previamente, existen varios modelos paramétricos que permiten describir los datos de supervivencia. Consecuentemente, es de gran utilidad contar con diagnósticos objetivos acerca de la conveniencia de una u otra distribución. Una de las alternativas consiste en graficar el  $\log(-\log(S_{KM}(t)))$  versus  $\log(\text{tiempo})$ , regresión que permite evaluar el comportamiento de la cola superior de la distribución y además, conducir a distribuciones que se ajustarán razonablemente en presencia de datos censurados (Cox, 1972).

En el mejoramiento genético animal, una de las formas paramétricas más utilizada para la función de riesgo base es el modelo Weibull, caracterizada por la simplicidad de sus funciones de riesgo, supervivencia y densidad. Para dicha distribución, y teniendo en cuenta la igualdad entre la función de riesgo y la función de supervivencia (expresión [3], sección 2.2.3), la función de riesgo base  $h_0(t) = \lambda\rho(\lambda t)^{\rho-1}$ , conduce a la función de supervivencia base  $S_0(t) = \exp\{-(\lambda t)^\rho\}$ . Asimismo, y definiendo la función de riesgo base acumulada para el modelo Weibull como  $H(t) = (\lambda t)^\rho$ , se tiene que:

$$\begin{aligned} \log[-\log S(t)] &= \log H(t) \\ &= \rho \log t + \rho \log \lambda \end{aligned} \quad [34]$$

Con lo cual, si los datos observados corresponden a una distribución Weibull, el resultado de graficar  $\log(-\log(S_{KM}(t)))$  versus  $\log(\text{tiempo})$  debe mostrar un ajuste lineal con pendiente  $\rho$  y ordenada al origen  $\rho \log \lambda$ . En este caso, el término  $S_{KM}(t)$  hace referencia al estimador Kaplan Meier de la función de supervivencia  $S(t)$  para toda la población.



**Figura 3.2.1.2.1.** Test gráfico para la distribución Weibull.

### 3.2.1.3. Significancia de efectos fijos - Prueba del cociente de verosimilitud (Kalbfleisch y Prentice, 1980).

Cualquiera sea el modelo ajustado (paramétrico o semi-paramétrico), resulta de interés verificar la significancia de los efectos incluidos en el modelo. En el análisis de supervivencia, al igual que en otras metodologías, la función de verosimilitud permite obtener estimaciones de los efectos incluidos en el modelo, testear su significancia y comparar los modelos ajustados. En este caso, el estimador  $\hat{\beta}$  del vector  $\beta$  puede calcularse maximizando la correspondiente función de verosimilitud. Dicho estimador tiene las mismas propiedades asintóticas que aquellas obtenidas de funciones regulares de la verosimilitud. En particular, el estimador es asintóticamente insesgado, y sigue una distribución normal multivariada con matriz de (co)varianzas igual a la inversa del Hessiano de la función de log-verosimilitud. Es decir que asintóticamente,  $\hat{\beta} \sim NMV\left(\beta, \left[i(\hat{\beta})\right]^{-1}\right)$ .

Esta propiedad asintótica hace factible el desarrollo de estadísticos que permitan poner a prueba la hipótesis  $\beta = \beta_0 = 0$ . No obstante, las diferencias en la función de verosimilitud cambian con la escala de medición, razón por la cual debe recurrirse a cocientes al momento de comparar modelos. Uno de los más conocidos es el *test del cociente de verosimilitud*, e igual a:

$$2\left\{\log\left(L(\beta_0)\right)-\log\left(L(\hat{\beta})\right)\right\} \quad [35]$$

En [35]  $\beta_0$  son los valores iniciales de los coeficientes y  $\hat{\beta}$  es la solución luego de ajustar el modelo. De esta forma, el test del cociente de verosimilitud compara los valores de la función de verosimilitud en  $\beta = 0$  y en  $\beta = \hat{\beta}$ , bajo  $H_0$ :

$$-2\log\left[\frac{\ell(\beta=0)}{\ell(\beta=\hat{\beta})}\right] = -2\log\left[\log\ell(\beta=0)-\log\ell(\beta=\hat{\beta})\right] \sim \text{asintóticamente } \chi_p^2 \quad [36]$$

Si se considera un modelo completo con parámetros de regresión  $\beta$  de orden  $p$ , y un modelo reducido con parámetros de regresión  $\beta^*$  de tamaño  $q$  ( $\beta^*$  incluye solo los primeros  $q$  elementos de  $\beta$ ), la  $H_0: \beta_{q+1} = \dots = \beta_p = 0$  se puede testear mediante:

$$-2 \log \left[ \log \ell(\beta = \hat{\beta}^*) - \log \ell(\beta = \hat{\beta}) \right] \quad [37]$$

El estadístico en [37] se distribuye asintóticamente como  $\chi^2_{p-q}$ .

#### 3.2.1.4. Extensión del modelo para incluir efectos aleatorios.

Con el fin de obtener estimaciones de las varianzas aditiva y entre-hatos, el modelo de efectos fijos, definido después de realizar el test del cociente de verosimilitud, puede ser ampliado de modo de incluir variables aleatorias como los efectos genéticos del animal, entre otras. Por lo tanto, para incorporar en el modelo el valor de cría de las hembras, sea  $v$  el término “frágil” o aleatorio no observado, componente que afecta multiplicativamente el riesgo de descarte de una vaca. En la evaluación genética animal, existe una relación entre  $v$  y el valor de cría del animal ( $a$ ), con  $a = \log v$ , y por lo tanto  $v = \exp \{a\}$  (Ducrocq y Casella, 1996). Esta última transformación permite incluir el efecto genético aditivo en la función de riesgo  $h(t)$ . Vaupel et al. (1979) indicaron que el término frágil hace referencia a las características genéticas no observables que afectan el riesgo de descarte.

De manera similar, se puede incorporar en el modelo el efecto aleatorio de hato. Así, la expresión de la función de riesgo de descarte o de concepción, para los caracteres vida productiva y días abiertos e intervalo primer-último servicio respectivamente, es la siguiente:

$$h(t) = h_0(t) \exp[\mathbf{x}'\beta + \mathbf{h} + \mathbf{a}] \quad [38]$$

El vector  $\mathbf{h}$  indica los efectos aleatorios de hato, y sigue una distribución normal con media cero y matriz de varianzas y covarianzas  $\mathbf{I} \sigma_h^2$ , y  $\mathbf{a}$  es el vector de valores de cría, que sigue una distribución normal multivariada con media cero y matriz de (co)varianzas  $\mathbf{A} \sigma_a^2$ , siendo  $\mathbf{A}$  la matriz de relaciones aditivas y  $\sigma_a^2$  la varianza genética aditiva. Para estimar las heredabilidades, se empleó el enfoque de Mészáros et al. (2010). En este caso, para un modelo animal, la heredabilidad está dada por:

$$h^2 = \frac{\sigma_a^2}{p^{-1} + \sigma_a^2 + \sigma_{hy}^2} \quad [39]$$

Donde el valor  $p$  corresponde a la proporción de animales no censurados para el carácter analizado.

#### 3.2.2. Modelo de datos agrupados

Como se mencionó anteriormente, el modelo de Cox carece de validez cuando el tiempo de falla es discreto con pocas categorías (es decir,  $< 20$ ), y existen muchas observaciones con el mismo tiempo de falla (empates). Así, sean los tiempos de falla agrupados en intervalos  $A_i = [a_{i-1}, \dots, a_i)$ ,  $i = 1, \dots, r$ , con  $a_0 = 0$  y  $a_r = \infty$ . Con lo cual, se tendrán intervalos que representan la unidad de medición (por ejemplo, años de vida

productiva):  $[a_1, a_2), \dots [a_{r-1}, a_r)$ . De acuerdo con Prentice y Gloeckler (1978), teniendo en cuenta los resultados obtenidos por Kalbfleisch y Prentice (1973), el vector de regresión puede ser tiempo-dependiente, pero fijo dentro de un intervalo específico de tiempo. Es decir,  $\mathbf{x} = \mathbf{x}(t_i)$  dentro del intervalo de tiempo  $A_i$ . Con lo cual, y retomando la función de supervivencia presentada en [3], para datos agrupados se tendrá:

$$S(t_{r-1}, \mathbf{x}) = \exp \left[ - \int_0^{t_{r-1}} h(u, \mathbf{x}) du \right] \quad [40]$$

De acuerdo con Lawless (1982), si se asume que el censurado ocurre solamente al final de cada intervalo, y si  $t = r_{i-1}$  representa el inicio del intervalo  $i$ , puede expresarse la expresión [40] del modo siguiente:

$$S(t_i, \mathbf{x}) = \prod_{j=1}^{i-1} \alpha_j^{\exp[\mathbf{x}(t_j)\boldsymbol{\beta}]} \quad [41]$$

Aquí, el término interno igual a

$$\alpha_j = \exp \left[ - \int_{a_{j-1}}^{a_j} h_0(u) du \right] \quad [42]$$

es la probabilidad de supervivencia condicional en  $A_j$  para un individuo con  $\mathbf{x}(t_j) = 0$ . En este caso, la función de riesgo es igual a:

$$h(t_i, \mathbf{x}) = \frac{S(a_{r-1}; \mathbf{x}) - S(a_r; \mathbf{x})}{S(a_{r-1}; \mathbf{x})} = 1 - \alpha_r^{\exp[\mathbf{x}(t_r)\boldsymbol{\beta}]} \quad [43]$$

Por su parte, la función de densidad  $f(t_i; \mathbf{x})$  estará dada por:

$$f(t_i, \mathbf{x}) = h(t_i, \mathbf{x}) S(t_i, \mathbf{x}) = \left( 1 - \alpha_r^{\exp[\mathbf{x}(t_r)\boldsymbol{\beta}]} \right) \left( \prod_{j=1}^{i-1} \alpha_j^{\exp[\mathbf{x}(t_j)\boldsymbol{\beta}]} \right) \quad [44]$$

Ahora, asumiendo que un tiempo de falla censurado  $t_k$  corresponde a un sujeto del cual se sabe que ha sobrevivido al inicio del intervalo  $A_k$ , y que  $\delta$  es el indicador de censura igual a 0 si la observación es censurada o 1 en caso contrario, al combinar [41] y [43] la función de verosimilitud es como sigue:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left( 1 - \alpha_k^{\exp[\mathbf{x}(t_k)\boldsymbol{\beta}]} \right)^\delta \prod_{j=1}^{k-1} \alpha_j^{\exp[\mathbf{x}(t_j)\boldsymbol{\beta}]} \quad [45]$$



### 3.2.2.1. Reparametrización del modelo de datos agrupados.

Por definición, los términos  $\alpha_i$  incluidos en [41] pueden tomar valores entre 0 y 1, requiriendo una maximización específica de [45]. Teniendo en cuenta el desarrollo presentado por Miller (1981), es conveniente reparametrizar los términos  $\alpha_i$  en  $\xi_i$ , donde  $\xi_i = \log(-\log \alpha_i)$ , los cuales pueden tomar valores entre  $-\infty$  y  $\infty$ . Entonces:

$$\alpha_i = \exp\{-\exp \xi_i\} \quad [46]$$

Esta reparametrización conduce a nuevas expresiones para las ecuaciones [41] y [43]

$$S(t, \mathbf{x}) = \exp\left\{-\mathbf{x}'\boldsymbol{\beta}\left(\exp^{\xi_1} + \exp^{\xi_2} + \dots + \exp^{\xi_{k-1}}\right)\right\} \quad [47]$$

y

$$h(t_i, \mathbf{x}) = 1 - \alpha_r^{\exp[\mathbf{x}'(t_j)\boldsymbol{\beta}]} = 1 - \exp\left\{-e^{\xi_k + \mathbf{x}'(t_j)\boldsymbol{\beta}}\right\} \quad [48]$$

En otras palabras, el modelo de Prentice and Gloecker (1978) toma forma paramétrica (exponencial:  $\rho = 1$ ), enfoque en el cual la función de riesgo base es estimada en todos los puntos de tiempo discretos, mediante el uso de una variable tiempo-dependiente adicional. Además, en el citado modelo la definición del riesgo se modifica para tener en cuenta la escala discreta de la variable tiempo (Ducrocq, 1999a).

### 3.2.2.2. Inclusión término frágil modelo de datos agrupados.

Bajo el enfoque de datos agrupados, la inclusión de un término frágil en el modelo es similar a la metodología presentada la sub-sección anterior. No obstante, y debido a que la reparametrización para contemplar tiempos de falla discretos se hace por medio de la función de supervivencia y de la función de riesgo, deben tenerse en cuenta las siguientes expresiones para la función de supervivencia:

$$S(t = \tau_{k-1} | w_m) = \exp\left(-\exp(w'_m \boldsymbol{\theta})\left(e^{\xi_1} + e^{\xi_2} + \dots + e^{\xi_{k-1}}\right)\right) \quad [49]$$

y para la función de riesgo:

$$h(t = \tau_{k-1} | w_m) = 1 - \exp\left[-e^{\xi_k + w'_m \boldsymbol{\theta}}\right] \quad [50]$$

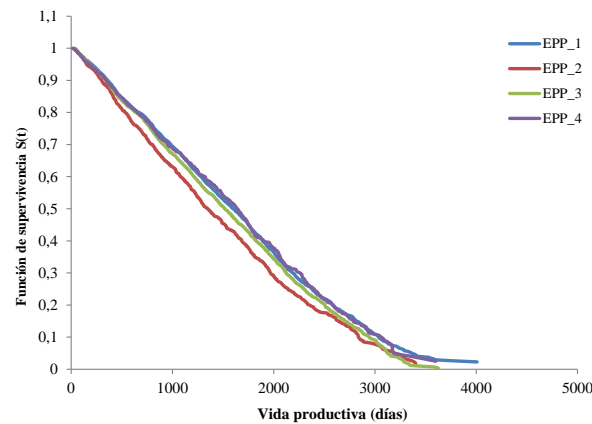
Dichas funciones deben ser contempladas al momento de construir la verosimilitud bajo un modelo de datos agrupados.

## ***Capítulo 4. Resultados.***

## Capítulo 4. Resultados.

### 4.1. Vida Productiva

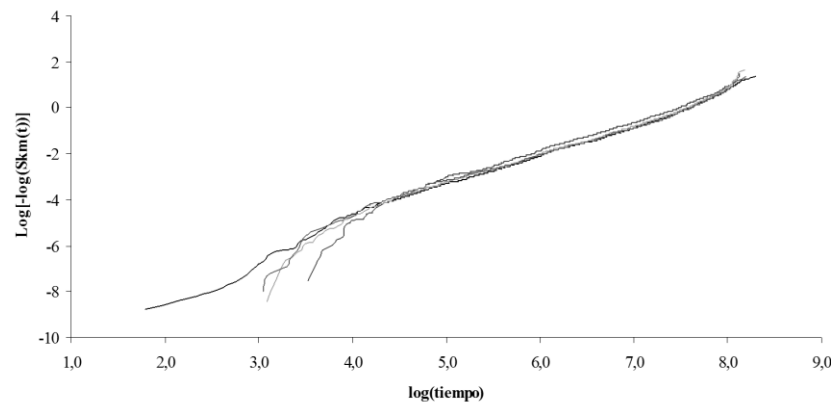
El porcentaje de registros censurados para este carácter fue de 51,8%. En primer lugar, se graficó la función de supervivencia para cada uno de los grupos de edad al primer parto, indicando que el 50 % de las vacas terminaron su vida productiva (VP) entre 1300 y 1700 días después del primer parto (Figura 4.1.1).



**Figura 4.1.1.** Función de supervivencia  $S(t)$  para vida productiva para los diferentes niveles de edad al primer parto.

#### 4.1.1. Control del supuesto de proporcionalidad en las funciones de riesgo.

La figura 4.1.1.1 muestra la variable  $\log(-\log(S_{KM,i}(t)))$  en función del  $\log(\text{tiempo})$  para los datos de VP, teniendo en cuenta en la construcción de estratos el efecto de edad al primer parto (EPP). En este caso,  $S_{KM,i}(t)$  hace referencia al estimador Kaplan-Meier o estimador del Producto Limite para la función de supervivencia  $S_i(t)$  en la  $i$ -ésima clase de EPP, y  $t$  es el número de días hasta el descarte del animal (Lawless, 1982). Si el supuesto de proporcionalidad se cumple, debe observarse paralelismo entre las líneas que representan cada una de las cuatro clases o estratos para EPP.

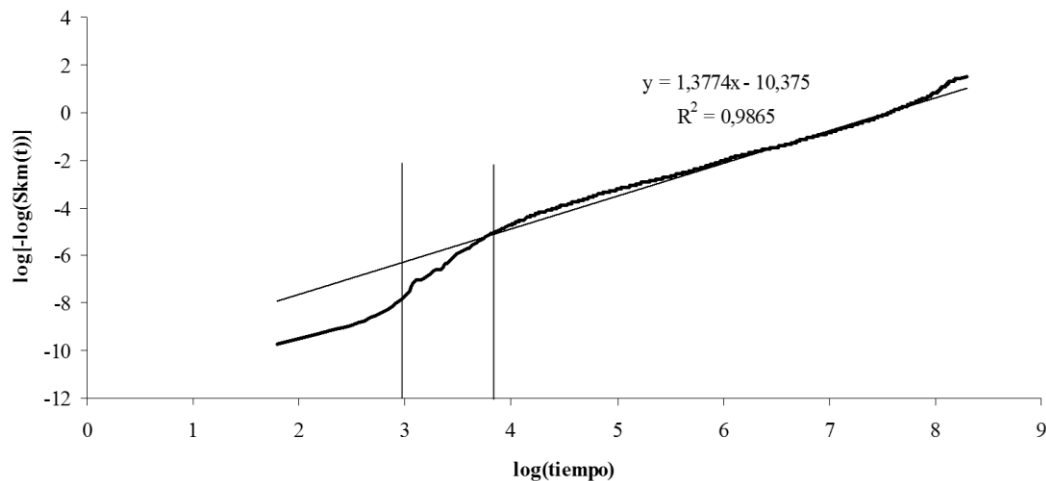


**Figura 4.1.1.1.** Supuesto de riesgo proporcional para vida productiva (VP).

En la figura 4.1.1.1, el supuesto de proporcionalidad se mantiene a través del tiempo principalmente después de  $\ln(t) = 4,22$ , es decir, después de 68 días aproximadamente. Aunque antes de dicho punto se observa disparidad en las líneas, el supuesto de proporcionalidad a través de grupos se mantiene en el tiempo. Por lo tanto, es adecuado ajustar un modelo de riesgo proporcional para VP.

#### 4.1.2. Control de la factibilidad del ajuste Weibull para la función de riesgo base

El estimador Kaplan - Meier de la función de supervivencia  $S_{KM}(t)$  también permite verificar, para el total de la población, el ajuste de un modelo de riesgo proporcional Weibull para la función de riesgo base. Si el supuesto es adecuado, la regresión de  $\log[-\log(S_{KM}(t))]$  versus  $\log(\text{tiempo})$  debe ser aproximadamente lineal.



**Figura 4.1.2.1.** Control del supuesto Weibull mediante regresión lineal de  $\text{Log}[-\log(S_{KM}(t))]$  versus  $\log(\text{tiempo})$  para vida productiva.

En la figura 4.1.2.1 se observan los siguientes intervalos en el ajuste: el primer período, comprendido entre el inicio del estudio y  $\ln(t) = 2,99$  (aproximadamente 20 días), es un ajuste lineal y paralelo a la recta de regresión; el segundo intervalo, comprendido entre  $\ln(t) = 2,99$  (20 días) y  $\ln(t) = 3,8$  (45 días) muestra un ajuste con una mayor pendiente que en el intervalo anterior; y el tercer período, comprendido entre  $\ln(t) = 3,8$  (45 días) y el final del período de análisis, sigue la tendencia descrita por la recta de regresión lineal. Con lo cual, y debido a la falta de ajuste Weibull en todo el período, es adecuado incluir una variable que represente el cambio en el riesgo de descartar a través del tiempo. De este modo, se definió el efecto fijo tiempo-dependiente del días post-primer parto, con tres niveles definidos por los intervalos [1;20 días), [7,5;45 días) y [45; 4540 días). Este efecto será incluido en el modelo de efectos fijos.

#### 4.1.3. Definición del modelo de efectos fijos.

Con el fin de determinar los factores que afectaron significativamente la vida productiva, se recurrió al test del cociente de verosimilitud, tanto para modelos que incluyen efectos fijos de forma sucesiva, como así también para comparar modelos que excluyan los efectos de forma individual (tablas 4.1.3.1. y 4.1.3.2.).

**Tabla 4.1.3.1.** Test de cociente de verosimilitud: factores incluidos secuencialmente.

Variable	GL Total	-2log(Verosimilitud)	Chi <sup>2</sup>	GL Delta	P-Value
None	2	129739,198			
NP	9	129013,559	725,64	7	0,0000
PL	13	126195,992	2817,6	4	0,0000
DPP	15	126107,21	88,782	2	0,0000
EPP	18	126098,212	8,9985	3	0,0293

None = Modelo sin efectos fijos; NP = Número de partos; PL = Producción de leche; DPP = Días post-primer parto; EPP = Edad al primer parto.

**Tabla 4.1.3.2.** Test del cociente de verosimilitud comparando el modelo completo vs. modelos excluyendo un efecto a la vez.

Variable	GL Total	-2log(Verosimilitud)	Chi <sup>2</sup>	GL Delta	P-Value
NP	11	126234,546	136,33	7	0,0000
PL	14	128901,24	2803	4	0,0000
DPP	16	126186,419	88,208	2	0,0000
EPP	15	126107,21	8,9985	3	0,0293

NP = Número de partos; PL = Producción de leche; DPP = Días post-primer parto; EPP = Edad al primer parto.

Tal como se observa en las tablas 4.1.3.1. y 4.1.3.2., todos los efectos fijos incluidos en el modelo fueron significativos a un nivel de 0,01 excepto la edad al primer parto (EPP) que fue significativa para  $\alpha = 0,05$ . Asimismo, se destaca el mayor impacto del número de partos (NP) y del nivel de producción de leche (PL) en el cambio del logaritmo de la verosimilitud, en comparación con los efectos de EPP y días post-primer parto (DPP). Por lo tanto, el modelo de riesgo proporcional Weibull para el carácter vida productiva fue:

$$h(t|\mathbf{x}'(t)) = \rho t^{\rho-1} \exp\left[\mathbf{x}'_{EPP}\boldsymbol{\beta}_{EPP} + \mathbf{x}'_{NP}(t)\boldsymbol{\beta}_{NP} + \mathbf{x}'_{PL}(t)\boldsymbol{\beta}_{PL} + \mathbf{x}'_{DPP}(t)\boldsymbol{\beta}_{DPP}\right] \quad [51]$$

donde:

- $\rho$  es el parámetro de forma que describe la función de riesgo base Weibull al igual que el parámetro  $\lambda$  de escala. No obstante, el parámetro  $\lambda$  es incluido en el término exponencial del modelo y representa la *media general*.
- *EPP* es el efecto fijo tiempo-independiente de edad al primer parto con 4 niveles: [18; 28 meses), [28; 32 meses), [32; 36 meses) y [36;42 meses).

- $NP(t)$ , es el efecto fijo tiempo-dependiente de número de parto, con  $\beta_{np}$  el vector de regresión relacionado a dicho efecto. El efecto de parto es considerado por incrementar el riesgo de descarte en vacas jóvenes. Se consideraron vacas con 1 a 12 partos, agrupando el número de partos en 1, 2, 3, 4, 5, 6, 7 y 8 o más partos. Hembras con más de 8 partos fueron consideradas censuradas en dicha parición.
- $PL(t)$  es el efecto fijo tiempo-dependiente de producción de leche, medido como desviaciones del promedio dentro de hato, con  $\beta_{prdl}$  el vector de regresión relacionado a dicho efecto. Se definieron cinco niveles:  $<80\%$ );  $[80;94\%)$ ,  $[94;104\%)$ ,  $[105;120\%)$  y  $[104;120\%)$ , y  $\geq 120\%$ . Las clases están basadas en los registros de lactancia ajustados a 305 días, con cambios en cada parición. Las vacas sin registros de producción en una lactancia en particular, fueron asignadas a la misma clase de producción que en la lactancia previa. El efecto de producción dentro de hato fue incluido para tener en cuenta el descarte voluntario debido a una baja producción. Los niveles fueron definidos teniendo en cuenta el número de animales promedio por hato (Vukasinovic et al., 2001).
- $DPP(t)$  es el efecto fijo tiempo dependiente, incluido para tener en cuenta el cambio en el riesgo de descarte a los 20 días y 45 días después del primer parto.

Para los efectos incluidos en el modelo [51], se estimaron los coeficientes de regresión, el error estándar, la tasa de riesgo (igual al término exponencial del estimador del coeficiente de regresión) y el número de observaciones no censuradas relacionadas con el error estándar del estimador, resultados que se presentan en la Tabla 4.1.3.3.

**Tabla 4.1.3.3.** Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), tasa de riesgo (TR), y número de observaciones no censuradas ( $n_{noc}$ ), para cada efecto incluido en el análisis de vida productiva.

Variable	Nivel	Est. Coef.	S.E	T.R.	$n_{nocens}$
NP	1	0,3661	0,1036	1,4420	2226
	2	0,4370	0,0834	1,5480	1598
	3	0,1149	0,0761	1,1220	1175
	4	0,0322	0,0729	1,0330	973
	5	0,0978	0,0734	1,1030	732
	6	0,0779	0,0778	1,0810	464
	7	0,0000	*	1,0000	260
	8	0,7250	0,1048	2,0650	141

Variable	Nivel	Est. Coef	S.E.	T.R.	n <sub>nocens</sub>
PL	1	1,8783	0,0431	6,5420	4421
	2	0,6653	0,0501	1,9450	898
	3	0,4536	0,0533	1,5740	687
	4	0,2503	0,0508	1,2840	842
	5	0,0000	*	1,0000	721
DPP	1	0,2423	0,3933	1,2740	7
	2	1,3620	0,1304	3,9040	88
	3	0,0000	*	1,0000	7474
EPP	1	0,0371	0,0342	1,0380	1221
	2	0,0000	*	1,0000	2902
	3	0,0074	0,0277	1,0070	2398
	4	0,1038	0,0361	1,1090	1048

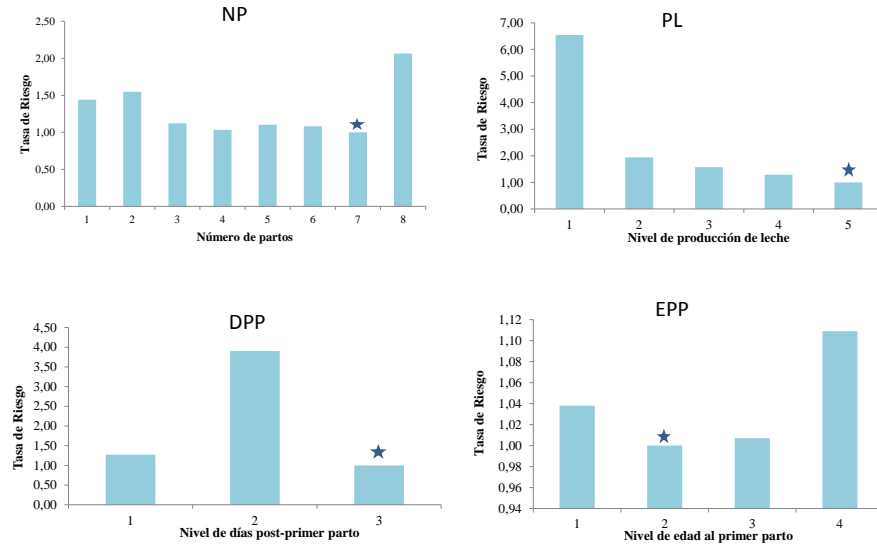
\* = Nivel de referencia para cada efecto (Fijado en 0, con TR igual a 1).

NP = Número de partos; PL = Producción de leche; DPP = Días post-primer parto; EPP = Edad al primer parto.

Para clarificar la relación entre los efectos del modelo y VP, es conveniente expresar los resultados en términos de *riesgo relativo* (RR). Este se define como el cociente entre la tasa de riesgo estimada para una vaca bajo la influencia de ciertos factores genéticos o ambientales y el riesgo medio, fijado en uno. La vaca de referencia tuvo su primer parto entre los 18 y 28 meses de edad, presentó 8 partos, y pertenece al nivel de mayor producción de leche y al último nivel de DPP.

Teniendo en cuenta los resultados presentados en la tabla 4.1.3.3 y en la figura 4.1.3.1, el riesgo relativo de descarte para vacas de 1 a 6 partos, fue 1,44, 1,55, 1,12, 1,03, 1,10 y 1,08 veces mayor en comparación con vacas que presentaron siete partos (nivel de referencia), es decir, hay un leve aumento en el riesgo entre el primer y el segundo parto, para luego mantenerse casi constante hasta la séptima parición. Sin embargo, el riesgo aumentó 2,07 veces en vacas que presentaron 8 o más partos en comparación con vacas pertenecientes al nivel de referencia. En relación al efecto PL, el riesgo relativo de descarte disminuye a medida que el nivel de producción aumenta, siendo 6,56 veces mayor en vacas correspondientes al nivel más bajo (nivel 1), al comparar con las vacas que pertenecen al nivel de producción más alto (nivel de referencia). A partir del nivel 1, el riesgo relativo de descarte disminuye drásticamente, siendo 1,95, 1,58 y 1,29 veces mayor para vacas clasificadas en los niveles 2, 3 y 4 de producción de leche respectivamente, en comparación con la de vacas en el nivel de referencia. El efecto DPP se caracterizó por un riesgo relativo

de descarte 1,27 y 3,9 veces mayor para vacas que se encontraban en los primeros 45 días después del primer parto en comparación con hembras con mayor número de días postparto. Finalmente, para EPP el riesgo relativo de descarte fue 1,04, 1,01 y 1,11 veces mayor en vacas pertenecientes a los niveles 1 (18-28 meses), 3 (32-36 meses) y 4 (36-42 meses) respectivamente, en comparación con el de hembras que presentaron su primer parto entre los 28 y 32 meses de edad (siendo el nivel 2 el grupo de referencia)



**Figura 4.1.3.1.** Tasa de riesgo para número de partos (NP), nivel de producción de leche (PL), efecto días post-primero parto (DPP), y edad al primer parto (EPP). (★) Nivel de referencia para cada efecto.

#### 4.1.4. Estimación de componentes de varianza y heredabilidad

Para estimar la varianza aditiva de las hembras para vida productiva, se incluyeron en el modelo [51] los efectos del animal. Asimismo, fue agregado el efecto de hato como aleatorio, de manera tal que:

$$h(t|x'(t)) = \rho t^{\rho-1} \exp \left[ x'_{EPP} \beta_{EPP} + x'_{NP}(t) \beta_{NP} + x'_{PL}(t) \beta_{PL} + x'_{DPP}(t) \beta_{DPP} + h + z' a \right] \quad [52]$$

Los nuevos efectos incluidos en [52] son los siguientes:

- ***h***, es el efecto aleatorio de hato, que sigue una distribución normal  $N(0, I\sigma_h^2)$ . La información de hato es incluida para tener en cuenta la influencia de las condiciones climáticas y ambientales del hato en la VP.
- ***a***, es el efecto aleatorio tiempo-independiente de animal, que sigue una distribución normal multivariada con vector de media  $\mathbf{0}$  y matriz de (co)varianza  $A\sigma_a^2$ , con ***A*** igual a la matriz de relaciones aditivas entre vacas.

El parámetro Weibull  $\rho$  para vida productiva fue estimado igual a 2,302, indicando un aumento en el riesgo de descarte a medida que pasa el tiempo. Asimismo, los valores estimados de varianza aditiva y de varianza de hato, con sus desvíos estándar, fueron



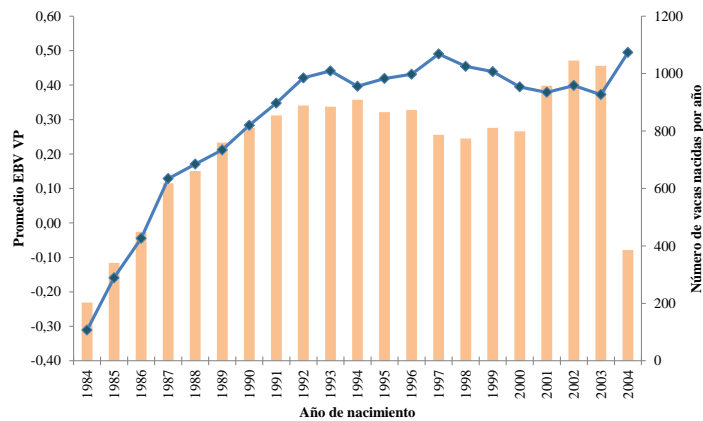
iguales a  $0,267 \pm 0,032$  y  $0,219 \pm 0,034$ , respectivamente. La heredabilidad bajo un modelo animal fue estimada teniendo en cuenta la siguiente expresión (Mészáros et al., 2010):

$$h^2 = \frac{\sigma_a^2}{p^{-1} + \sigma_a^2 + \sigma_h^2} \quad [53]$$

donde  $\sigma_a^2$  es la varianza genética,  $\sigma_h^2$  es la varianza de hato y  $p$  la proporción de animales no censurados. En este caso, el efecto aditivo animal explica toda la varianza genética, para lo cual son incluidos todos los individuos relacionados en el pedigree. Para este carácter  $p = 0,482$ , con lo cual la heredabilidad para vida productiva (con su desvío estándar) fue igual a  $0,104 \pm 0,011$ .

#### 4.1.5. Tendencia genética.

La figura 4.1.5.1 muestra la tendencia genética para vida productiva en el periodo 1984-2004, así como el número de hembras nacidas en cada año. En términos generales, y de acuerdo a la figura, la tendencia genética para VP funcional es ascendente excepto por el período comprendido entre 1997 y 2003. Sin embargo, y dado que los valores de cría están expresados como tasas de riesgo relativas de descarte, la tendencia genética representa un comportamiento desfavorable de los valores de cría: cualquiera haya sido el criterio de selección, tuvo un efecto negativo en el valor de cría para VP de las vacas nacidas entre 1984 y 1997. A partir de dicho punto y hasta 2003, la tendencia genética disminuyó dando lugar a un aumento en los valores de cría promedio para VP. Sin embargo, esta tendencia cambió en el año 2004.



**Figura 4.1.5.1.** Tendencia genética para vida productiva. Barras verticales = número de vacas nacidas por año; línea = tendencia para vida productiva.

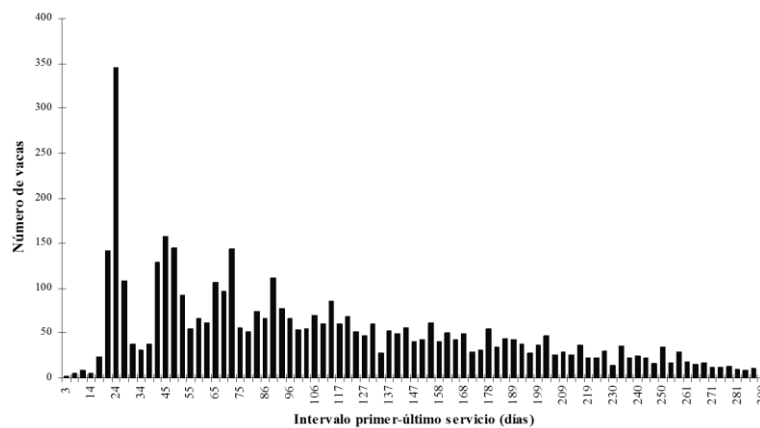
#### 4.1.6. Correlación genética para vida productiva y producción de leche.

Siguiendo el enfoque de Calo et al. (1973), se estimó la correlación genética entre los valores de cría para vida productiva y aquellos obtenidos para producción de leche. Estos últimos fueron estimados ajustando un modelo lineal mixto incluyendo los efectos fijos de hato, año de parto, mes de parto, edad al primer parto, y efectos aleatorios de ambiente permanente para cada registro de producción, de animal y el efecto residual. En este caso,  $pe \sim N(0, \mathbf{I} \sigma_p^2)$ ,  $a \sim N(0, \mathbf{A} \sigma_a^2)$  y  $e \sim N(0, \mathbf{I} \sigma_e^2)$  donde  $\mathbf{A}$  es la matriz de relaciones de parentesco.

En la estimación se tuvo en cuenta la corrección por error de muestreo. De este modo, la correlación genética estimada entre los valores de cría para VP y PL fue de -0.3. No obstante, y dado que los valores de cría para VP están expresados como tasas de riesgo relativas de descarte, el valor obtenido de correlación implicaría que a mayor nivel de producción de leche, se tendría una disminución en los valores de cría para riesgo de descarte y por tanto, aumentos en la vida productiva de las hembras evaluadas.

## 4.2. Intervalo primer-último servicio

El carácter de intervalo entre el primer y el último servicio (IPU) es definido como el número de días entre el primer servicio después del último parto y el último servicio, o el diagnóstico de preñez. Al igual que para VP, IPU tiene una distribución sesgada a la derecha (Figura 4.2.1).



**Figura 4.2.1.** Distribución de frecuencias para IPU, excluyendo vacas que consiguieron una preñez en la primera inseminación (39,38% de las vacas).

Sin embargo, y a diferencia de VP, IPU tiene la particularidad de presentar varias observaciones repetidas o *empates*, aspecto que sugiere el ajuste de un modelo de datos agrupados como sugirieron Prentice y Gloeckler (1978). De esta manera, el carácter IPU medido inicialmente en días, fue discretizado teniendo en cuenta la distribución que se muestra en la figura 5, dando lugar a 15 períodos definidos así: día 1; día 2-11; y luego, cambios cada 21 días que representan la primera, segunda y posteriores inseminaciones o servicios, coincidiendo con la duración promedio del ciclo estral en bovinos: día 12-33; día 34-55; día 56-77; día 78-99; día 100-121; día 122-143; día 144-165; día 166-187; día 188-209; día 210-231; día 232-253; día 254-275; y  $\geq 276$  días. Así, el porcentaje de observaciones censuradas fue igual a 56,5 %.

### 4.2.1. Modelo de datos agrupados

Teniendo en cuenta la discretización realizada para el carácter IPU, se ajustó el siguiente modelo de datos agrupados (Prentice y Gloeckler, 1978):

$$h(t|x(t)) = 1 - \exp \left[ -e^{\xi_k + x'_{EPP} \beta_{EPP} + x'_{NP} \beta_{NP} + x'_{SC} \beta_{SC} + x'_{TU}(t) \beta_{TU}} \right] \quad [54]$$

donde:

- $\xi_k = \log(-\log \alpha_k)$  es la reparametrización de los  $\alpha_i$ 's definidos en [46], que permite que tomen valores entre  $-\infty$  y  $+\infty$ ; y  $\alpha_k = \exp\left\{-\int_{\tau_{i-1}}^{\tau_i} h_0(u) du\right\}$  es la función de supervivencia base en el intervalo  $[\tau_{i-1}, \tau_i)$ .
- *EPP* es el efecto fijo tiempo-independiente de edad al primer parto con 4 niveles: [18; 28 meses), [28; 32 meses), [32; 36 meses) y [36;42 meses).
- *NP* es el efecto fijo tiempo-independiente de número de partos previos a la medición de IPU, con  $\beta_{NP}$  como el vector de regresión relacionado con dicho efecto. El efecto de número de partos es afectar el retorno a la ciclicidad por parte de la hembra. Se consideraron vacas con 1 a 6 partos.
- *SC* es el efecto fijo tiempo-independiente de número de servicios por ciclo, con  $\beta_{SC}$  el vector de regresión relacionado a dicho efecto, medido como variable continua o covariable. El efecto se obtuvo dividiendo el número de inseminaciones que recibió la vaca, por el número de ciclos estrales transcurridos entre el primer y el último servicio o diagnóstico de preñez. El objetivo de este efecto es evaluar la eficiencia, tanto para la detección de celos como para la calidad del proceso de inseminación.
- *TU*, es la covariable tiempo dependiente generada de forma automática en el Survival Kit v6.0 (Time\_unit; Ducrocq et al., 2010) al ajustar un modelo de datos agrupados. En este caso, la covariable *TU* toma los valores 1, 2, 3,..., 15, correspondientes al número de ciclos entre el primer servicio y el diagnóstico de preñez o último servicio (tiempo de censura). El propósito de esta covariable es estimar conjuntamente, tanto la función de riesgo, como los parámetros de regresión relacionados con los efectos fijos incluidos en el modelo.

Para evaluar la significancia de los factores que afectan el riesgo de concepción en las hembras, en las tablas 4.2.1.1 y 4.2.1.2 se observan los valores del test del cociente de verosimilitud para modelos que incluyen los efectos fijos de forma secuencial, como para aquellos modelos donde se excluyeron los efectos individuales.

**Tabla 4.2.1.1.** Test del cociente de verosimilitud, incluyendo factores secuencialmente al modelo de efectos fijos.

Variable	GL Total	-2log(Verosimilitud)	Chi <sup>2</sup>	GL Delta	P-Value
None	1	19139,6358	-	-	-
NP	6	19053,7148	85,921	5	<0,001
SC	7	16378,9999	2674,7	1	<0,001
EPP	10	16374,2264	4,7735	3	0,1892
TU	25	15512,2211	862,01	15	<0,001

None = Modelo sin inclusión de efectos fijos; NP = Número de partos; SC = Servicios por ciclo estral; EPP = Edad al primer parto; TU = Variable Time\_unit.

**Tabla 4.2.1.2.** Test del cociente de verosimilitud comparando el modelo completo contra aquellos excluyendo de un efecto a la vez.

Variable	GL Total	-2log(Verosimilitud)	$\chi^2$	gl Delta	P-Value
NP	20	15590,2563	78,035	5	<0,001
SC	25	15512,22	0,000	0	<0,001
EPP	22	15518,0731	5,852	3	0,1190
TU	10	16374,2264	862,01	15	<0,001

NP = Número de partos; SC = Servicios por ciclo estral; EPP = Edad al primer parto; TU = Variable Time\_unit.

De acuerdo con las tablas 4.2.1.1 y 4.2.1.2, todos los factores fueron significativos en el análisis de IPU excepto el efecto de edad al primer parto. Con lo cual, y excluyendo dicho efecto, el modelo ajustado para el carácter IPU fue:

$$h(t|x(t)) = 1 - \exp \left[ -e^{\xi_k + x'_{NP} \beta_{NP} + x'_{SC} \beta_{SC} + x'_{TU}(t) \beta_{TU}} \right] \quad [55]$$

siendo los componentes del modelo definidos al igual que se hizo en [54], pero excluyendo el efecto de edad al primer parto. Al ajustar el modelo dado en la expresión [55], se obtuvieron los estimadores de los coeficientes de regresión, el error estándar, la tasa de riesgo para cada uno de los niveles de los efectos incluidos, así como el número de observaciones no censuradas, información que se presenta en la tabla 4.2.1.3.

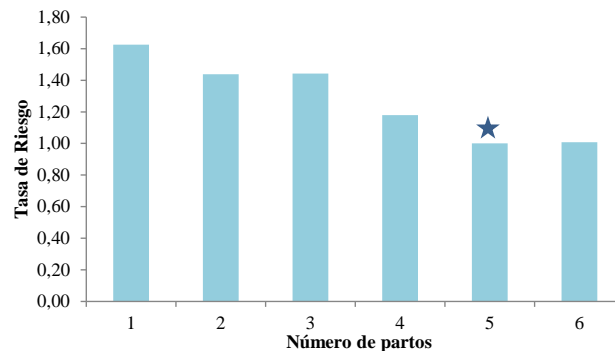
**Tabla 4.2.1.3.** Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), Tasa de riesgo (TR), y número de observaciones no censuradas ( $n_{nocens}$ ), para cada efecto incluido en el análisis del carácter IPU.

Efecto	Nivel	Est.Coef	S.E.	T.R.	$n_{nocens}$
NP	1	0,4854	0,076	1,6250	703
	2	0,3631	0,0755	1,4380	746
	3	0,3665	0,0786	1,4430	548
	4	0,1651	0,0852	1,1800	341
	5	0,0000	*	1,0000	233
	6	0,0071	0,0912	1,0070	251
SC	1	5,071	0,1231	-	2822

\* = Nivel de referencia para cada efecto (Fijado en 0, con TR igual a 1).

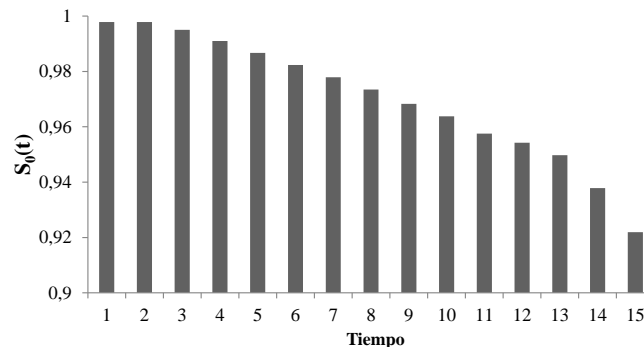
NP = Número de partos; SC = Servicios por ciclo estral.

Teniendo en cuenta los resultados presentados en la tabla 4.2.1.3 y en la figura 4.2.1.1, las vacas de primera parición tuvieron un riesgo de concepción 1,63 veces mayor que vacas de cinco partos (nivel de referencia). Este riesgo disminuyó para las demás categorías, siendo el riesgo de concepción 1,44 veces mayor para vacas que presentaron dos y tres partos en comparación con las de quinta parición. Asimismo, vacas con 4 y 6 partos tuvieron un riesgo de concepción 1,18 y 1,01 veces mayor que las hembras pertenecientes al nivel de referencia.



**Figura 4.2.1.1.** Tasa de riesgo relativo para número de partos. ( ★ Nivel de referencia.

Asimismo, en la figura 4.2.1.2 se presenta el comportamiento de la función base de supervivencia estimada para cada uno de los tiempos discretos de la variable IPU. Debido a la discretización, se tiene una función base de supervivencia segmentada, dando lugar a una función de supervivencia por partes (“stepwise function”) y por lo tanto, a una función de riesgo particular para cada uno de los tiempos discretizados.



**Figura 4.2.1.2.** Estimación discreta de la función base de supervivencia para el carácter intervalo primer-último servicio.

**Tabla 4.2.1.4.** Estimación discreta de la función base de supervivencia para el intervalo primer-último servicio.

T	Est.	S.E.	S(t)Cond	L.Inf S(t)	L.Sup S(t)	S <sub>0</sub> (t)	n <sub>nocens</sub>
1	-6,1603	0,0759	0,9979	0,9976	0,9982	0,997	1229
2	-9,9878	0,4475	1,0000	0,9999	1,0000	0,9978	5
3	-5,873	0,0873	0,9972	0,9967	0,9976	0,9950	283
4	-5,5019	0,0864	0,9959	0,9952	0,9966	0,9910	296
5	-5,45	0,0913	0,9957	0,9949	0,9964	0,9867	235
6	-5,4078	0,097	0,9955	0,9946	0,9963	0,9823	189
7	-5,3975	0,1034	0,9955	0,9945	0,9963	0,9779	151
8	-5,3929	0,1119	0,9955	0,9944	0,9964	0,9735	117
9	-5,2318	0,1156	0,9947	0,9933	0,9958	0,9683	107
10	-5,3738	0,137	0,9954	0,9940	0,9965	0,9638	68
11	-5,0331	0,1385	0,9935	0,9915	0,9950	0,9575	67
12	-5,6849	0,2147	0,9966	0,9948	0,9978	0,9543	24
13	-5,3526	0,2279	0,9953	0,9926	0,9970	0,9498	21
14	-4,3701	0,2193	0,9874	0,9807	0,9918	0,9378	23
15	-4,0671	0,3845	0,9830	0,9643	0,9920	0,9219	7

T= Tiempo; Est = Estimador para cada uno de los tiempos; S.E: error estándar; S(t) Cond = Función de supervivencia condicional (desde  $t < 1$  hasta  $t$ ); L.InfS(t) y L.SupS(t) = Límites inferior y superior para S(t); S<sub>0</sub>(t) = Función de supervivencia base resultante; n<sub>nocens</sub> = número de observaciones no censuradas.

#### 4.2.2. Estimación de componentes de varianza y heredabilidad

A los efectos de estimar la varianza aditiva de las hembras para el carácter intervalo entre primer y último servicio, se agregaron los efectos del animal al modelo [55]. Asimismo, se incluyó el efecto de hato como aleatorio, de modo tal que

$$h(t|x(t)) = 1 - \exp \left[ -e^{\xi_k + \mathbf{x}'_{NP} \boldsymbol{\beta}_{NP} + \mathbf{x}'_{SC} \boldsymbol{\beta}_{SC} + \mathbf{x}'_{TU}(t) \boldsymbol{\beta}_{TU} + \mathbf{h} + \mathbf{z}' \mathbf{a}} \right] \quad [56]$$

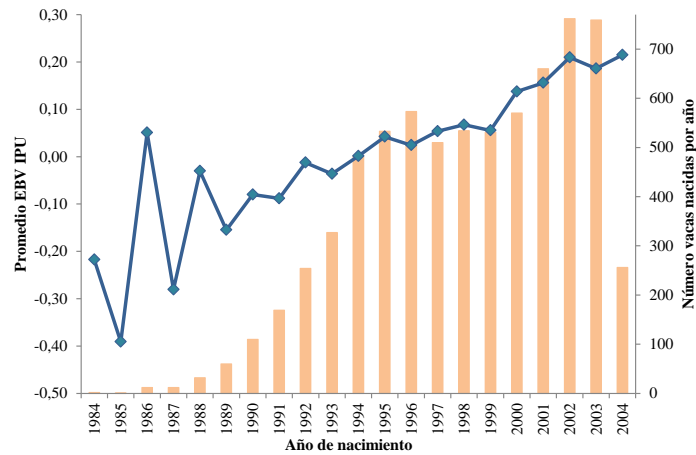
donde

- $h$ , es el efecto aleatorio de hato, que sigue una distribución normal  $N(0, I\sigma^2_h)$ . La información de hato es incluida para tener en cuenta la influencia de las condiciones climáticas y ambientales del hato en el IPU.
- $a$ , es el efecto aleatorio tiempo-independiente de animal, de distribución normal multivariada con vector de media  $\mathbf{0}$  y matriz de (co)varianza  $A\sigma^2_a$ ;  $A$  es la matriz de relaciones aditivas entre vacas.

Para el carácter IPU, los valores estimados de varianza aditiva y de varianza de hato, así como los valores de desvío estándar fueron 0,3980 (0,02) y 0,999 (0,000) respectivamente. Nuevamente, la heredabilidad en el modelo animal fue estimada teniendo en cuenta la expresión [53]. Dado que el porcentaje de censura fue 60,6 %, la proporción de registros no censurados  $p$  es igual 0,3946. Así, la heredabilidad para el intervalo entre el primer y el último servicio (con su desvío estándar) fue  $0,1013 \pm 0,0046$ .

#### 4.2.3. Tendencia genética

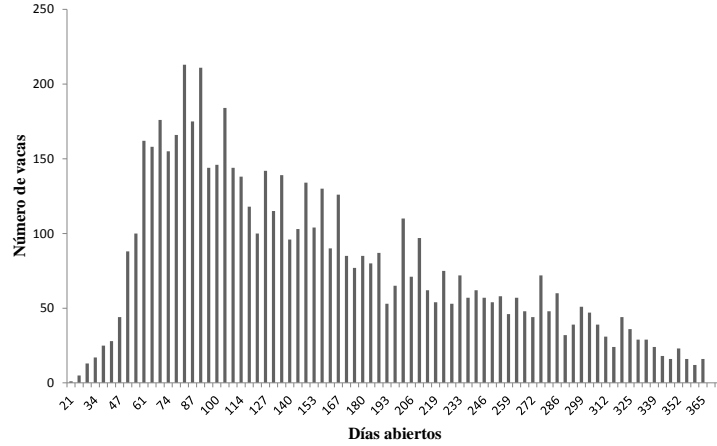
La figura 4.2.3.1 muestra la tendencia genética para el carácter IPU, así como el número de vacas que nacieron en el período 1984 – 2004 y que fueron evaluadas para dicho carácter. Se puede observar que la tendencia genética es ascendente. Dado que los valores de cría están expresados como tasas de riesgo al igual que para el carácter VP, y más precisamente como riesgo de concepción, la tendencia genética representa un comportamiento favorable de los valores de cría de IPU. Es decir, los valores de cría para el riesgo de concepción aumentaron, dando lugar a menores intervalos entre el primer y el último servicio.



**Figura 4.2.3.1.** Tendencia genética para intervalo primer-último servicio. Barras verticales = número de vacas nacidas por año; línea = tendencia genética para IPU.

#### 4.3. Días abiertos

Debido al comportamiento sesgado del carácter días abiertos (DA), es razonable asumir una distribución Weibull para la función de riesgo base (Figura 4.3.1).



**Figura 4.3.1.** Distribución de frecuencias para el carácter días abiertos.

Al igual que IPU, DA tiene la particularidad de presentar *empates*, sugiriendo nuevamente el ajuste de un modelo de datos agrupados (Prentice y Gloeckler, 1978). De este modo, el carácter DA medido en días, fue discretizado teniendo en cuenta en este caso, la distribución observada en la figura 4.3.1. El carácter DA fue dividido en 16 grupos, cambiando cada 21 días, período que representa la duración promedio del ciclo estral en bovinos. Estos períodos en días fueron: 21-44; 45-66; 67-88, ..., y  $\geq 353$  días.

#### 4.3.1. Modelo de datos agrupados

Teniendo en cuenta la discretización realizada para DA, se ajustó un modelo de datos agrupados siguiendo el enfoque de Prentice y Gloeckler (1978), dado por:

$$h(t|x(t)) = 1 - \exp \left[ -e^{\xi_k + \mathbf{x}'_{EPP} \boldsymbol{\beta}_{EPP} + \mathbf{x}'_{NP} \boldsymbol{\beta}_{NP} + \mathbf{x}'_{SC} \boldsymbol{\beta}_{SC} + \mathbf{x}'_{TU}(t) \boldsymbol{\beta}_{TU}} \right] \quad [57]$$

donde:

- $\xi_k = \log(-\log \alpha_k)$  es la reparametrización de los  $\alpha_i$ 's definidos en [46], que permite que tomen valores entre  $-\infty$  y  $+\infty$ ; y  $\alpha_k = \exp \left\{ -\int_{\tau_{i-1}}^{\tau_i} h_0(u) du \right\}$  es la función de supervivencia base en el intervalo  $[\tau_{i-1}, \tau_i)$ .
- *EPP* es el efecto fijo tiempo-independiente de edad al primer parto con 4 niveles: [18, 28 meses); [28, 32 meses); [32, 36 meses); y [36, 42 meses).
- *NP*, es el efecto fijo tiempo-independiente de número de partos (1 a 6 pariciones) previos a la medición de DA, considerado por afectar el retorno a la ciclicidad en la hembra, y con  $\beta_{Npart}$  el vector de regresión relacionado a dicho efecto.
- *SC* es el efecto fijo tiempo-independiente de número de servicios por ciclo, medido como covariable continua y con  $\beta_{servic}$  como vector de regresión. El efecto fue construido dividiendo el número de inseminaciones por el número de ciclos estrales transcurridos entre el parto y el último servicio o diagnóstico de preñez.
- *TU*, es la covariable tiempo dependiente generada de forma automática en el Survival Kit v6.0 (Ducrocq et al., 2010). En este caso, la covariable time\_unit toma los valores 1,



2, 3,..., 16, correspondientes al número de ciclos entre el parto y el diagnóstico de preñez o último servicio (tiempo de censura). Al igual que en el carácter IPU, esta covariable permite estimar conjuntamente tanto la función de riesgo como los parámetros de regresión.

Las tablas 4.3.1.1 y 4.3.1.2 muestran las pruebas del cociente de verosimilitud para DA, incluyendo y excluyendo efectos fijos, de uno a la vez. El objetivo fue evaluar la significancia de los factores que afectan el riesgo de concepción en las hembras.

**Tabla 4.3.1.1.** Test del cociente de verosimilitud, incluyendo factores de forma secuencial.

Variable	GL Total	-2log(likelihood)	$\chi^2$	gl Delta	P-Value
None	1	19810,9696			
NP	6	19769,6476	41,322	5	0,0000
SC	7	19659,6592	109,99	1	0,0000
EPP	10	19643,6724	15,987	3	0,0011
TU	26	18699,335	944,34	16	0,0000

None = Modelo sin inclusión de efectos fijos; NP = Número de partos; SC = Servicios por ciclo estral; EPP = Edad al primer parto; TU = Variable Time\_unit.

**Tabla 4.3.1.2.** Test del cociente de verosimilitud comparando el modelo completo vs. modelos excluyendo un efecto a la vez.

Variable	GL Total	-2log(likelihood)	$\chi^2$	gl Delta	P-Value
NP	21	18746,3495	47,015	5	0,0000
SC	26	18699,3382	0,000	0	0,0000
EPP	23	18719,7649	20,43	3	0,0001
TU	10	19643,6724	944,34	16	0,0000

NP = Número de partos; SC = Servicios por ciclo estral; EPP = Edad al primer parto; TU = Variable Time\_unit.

El porcentaje de observaciones censuradas para DA fue de 56,5%. Como se observa en las tablas 4.3.1.1 y 4.3.1.2, todos los efectos fijos resultaron significativos, sea en los modelos donde los efectos fijos se tomaron en forma secuencial como individual. Los estimadores de los coeficientes de regresión, el error estándar, la tasa de riesgo y el número de observaciones no censuradas en cada uno de los niveles de los efectos fijos se presentan en la Tabla 4.3.1.3.

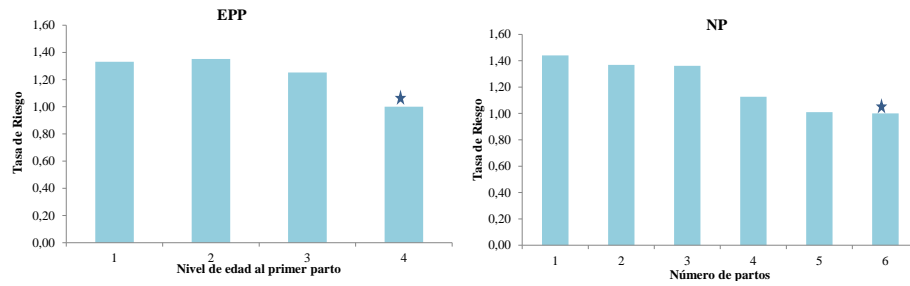
**Tabla 4.3.1.3.** Estimaciones para el coeficiente de riesgo (Est. Coef.), error estándar (S.E.), Tasa de riesgo (TR) y número de observaciones no censuradas ( $n_{\text{nocens}}$ ), para cada nivel de los efectos incluidos en el análisis de días abiertos.

<b>Efecto</b>	<b>Nivel</b>	<b>Est.Coef</b>	<b>S.E.</b>	<b>T.R.</b>	<b><math>n_{\text{nocens}}</math></b>
NP	1	0,3645	0,0785	1,440	687
	2	0,3133	0,0782	1,368	714
	3	0,3088	0,0811	1,362	532
	4	0,1197	0,0879	1,127	331
	5	0,0100	0,0960	1,010	221
	6	0,0000	*	1,000	214
SC	1	1,7330	0,0792	*	2699
EPP	1	0,2858	0,0766	1,331	624
	2	0,3013	0,0711	1,352	1236
	3	0,2251	0,0769	1,252	603
	4	0,0000	*	1,000	236

\* = Nivel de referencia para cada efecto (Fijado en 0, con TR igual a 1).

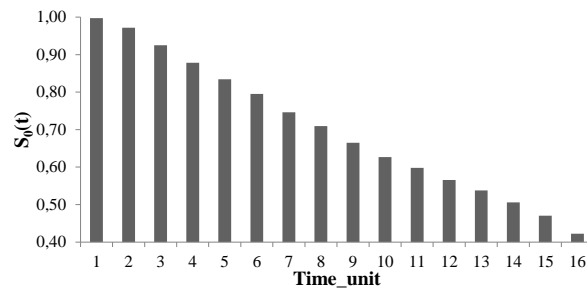
NP = Número de partos; SC = Servicios por ciclo estral; EPP = Edad al primer parto.

A partir de los resultados en la tabla 4.3.1.3 y la figura 4.3.1.1, se puede indicar que vacas con *EPP* entre 18 y 28 meses (nivel 1), 28 y 32 meses (nivel 2) y entre 32 y 36 meses (nivel 3), tuvieron un riesgo relativo de concepción 1,33, 1,35 y 1,25 veces mayor que aquellas hembras que presentaron su primer parto entre los 36 y 42 meses de edad (nivel 4 – grupo de referencia). Es decir, el riesgo de conseguir una preñez disminuyó a medida que la *EPP* fue mayor. En el caso del efecto NP, el riesgo relativo de concepción mostró una tendencia decreciente siendo 1,44, 1,13 y 1,01 veces mayor para vacas de 1, 4 o 5 partos, respectivamente, en relación con los de vacas de 6 años (nivel de referencia).



**Figura 4.3.1.1.** Tasa de riesgo para edad al primer parto (EPP) y número de partos (NP). (★) Nivel de referencia para cada efecto.

La figura 4.3.1.2 muestra la función base de supervivencia, estimada para cada uno de los tiempos discretos de la variable DA. Al igual que para el caracter, la función de supervivencia es segmentada y produce a una curva de supervivencia por partes (“stepwise function”) debido a la discretización del caracter.



**Figura 4.3.1.2.** Estimación discreta de la función base de supervivencia para DA.

**Tabla 4.3.1.4.** Estimación discreta de la función base de supervivencia para el carácter días abiertos.

T	Est.	S.E.	S(t)Cond	L.Inf S(t)	L.Sup S(t)	S <sub>0</sub> (t)	N <sub>nocens</sub>
1	-5,9562	0,1958	0,9974	0,9962	0,9982	0,9974	37
2	-3,6482	0,109	0,9743	0,9683	0,9792	0,9718	285
3	-3,0156	0,1025	0,9522	0,9418	0,9607	0,9253	453
4	-2,9569	0,1039	0,9494	0,9383	0,9585	0,8784	390
5	-2,9651	0,1067	0,9498	0,9384	0,9590	0,8343	315
6	-3,0414	0,1107	0,9534	0,9424	0,9623	0,7954	243
7	-2,7521	0,1091	0,9382	0,9240	0,9498	0,7462	262
8	-2,9898	0,1183	0,9510	0,9386	0,9609	0,7096	168
9	-2,7288	0,1171	0,9368	0,9211	0,9494	0,6648	177
10	-2,828	0,1265	0,9426	0,9270	0,9549	0,6266	126
11	-3,0557	0,1444	0,9540	0,9394	0,9651	0,5978	79
12	-2,8893	0,1509	0,9459	0,9280	0,9595	0,5654	69
13	-2,9905	0,1806	0,9510	0,9309	0,9653	0,5377	41
14	-2,7916	0,2038	0,9405	0,9126	0,9597	0,5057	30
15	-2,6179	0,2587	0,9296	0,8859	0,9570	0,4702	17
16	-2,2355	0,3892	0,8986	0,7951	0,9514	0,4225	7

T= Tiempo; Est.= Estimador para cada uno de los tiempos; S.E.: error estándar; S(t)Cond = Función de supervivencia condicional (desde  $t < 1$  hasta  $t$ ); L.Inf S(t) y L.Sup S(t) = Límites inferior y superior para S(t); S<sub>0</sub>(t) = Función de supervivencia base resultante; n<sub>nocens</sub> = número de observaciones no censuradas.

#### 4.3.2. Estimación de los componentes de varianza y la heredabilidad

La varianza aditiva para DA fue estimada empleando el siguiente modelo:

$$h(t|x(t)) = 1 - \exp \left[ -e^{\xi_k + \mathbf{x}'_{EPP} \boldsymbol{\beta}_{EPP} + \mathbf{x}'_{NP} \boldsymbol{\beta}_{NP} + \mathbf{x}'_{SC} \boldsymbol{\beta}_{SC} + \mathbf{x}'_{TU}(t) \boldsymbol{\beta}_{TU} + \mathbf{h} + \mathbf{z}' \mathbf{a}} \right] \quad [58]$$

donde

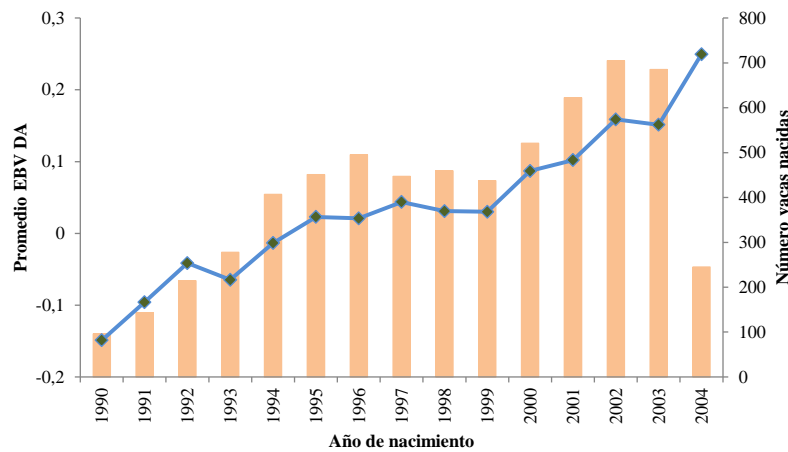
- $\mathbf{h}$  es el efecto aleatorio de hato, que sigue una distribución normal,  $\mathbf{h} \sim \mathbf{N}(0, \mathbf{I}\sigma_h^2)$ .

- $a$  es el efecto aleatorio tiempo-independiente de animal, que sigue una distribución normal multivariada con vector de media  $\mathbf{0}$  y matriz de (co)varianzas  $\mathbf{A}\sigma_a^2$ , siendo  $\mathbf{A}$  la matriz de relaciones aditivas entre vacas.

Las estimaciones de la varianza aditiva y la de hato, con sus respectivos desvíos estándar asociados fueron  $0,245 \pm 0,074$  y  $0,292 \pm 0,061$ . La heredabilidad de un modelo animal fue nuevamente estimada a través de la expresión [53]. La proporción de animales no censurados ( $p$ , Mészáros et al., 2010) fue igual a 0,435. Consecuentemente, la heredabilidad de DA, junto con su desvío estándar, fue igual a  $0,086 \pm 0,024$ .

#### 4.3.3. Tendencia genética.

La figura 4.3.3.1 muestra la tendencia genética para DA, junto con el número de vacas nacidas por año durante el período 1990-2004. Al igual que para caracteres anteriores, la tendencia genética expresada en términos de tasas de riesgo, muestra un deterioro de los valores de cría predichos de DA con el tiempo.



**Figura 4.3.3.1.** Tendencia genética para días abiertos. Barras verticales representan el número de vacas nacidas por año; la línea representa tendencia para días abiertos.

#### 4.3.5. Correlación genética para días abiertos y producción de leche.

Siguiendo nuevamente el enfoque de Calo et al. (1973), se estimó la correlación genética entre los valores de cría para días abiertos y aquellos obtenidos para producción de leche. Estos últimos fueron estimados ajustando un modelo lineal mixto incluyendo los efectos fijos de hato, año de parto, mes de parto, edad al primer parto, y efectos aleatorios de ambiente permanente para cada registro de producción, de animal y el efecto residual. En este caso,  $pe \sim N(0, \mathbf{I}\sigma_p^2)$ ,  $a \sim N(0, \mathbf{A}\sigma_a^2)$  y  $e \sim N(0, \mathbf{I}\sigma_e^2)$  donde  $\mathbf{A}$  es la matriz de relaciones de parentesco. En la estimación se tuvo en cuenta la corrección por error de muestreo. De este modo, la correlación genética estimada entre los valores de cría para DA y PL fue de 0.47. Ahora bien, dado que los valores de cría para DA están expresados como tasas de riesgo relativas de concepción, el valor obtenido de correlación implicaría que a mayor nivel de

producción de leche, se tendría un aumento en los valores de cría para riesgo de concepción y por tanto, disminución en los días abiertos para las hembras bajo estudio.

## ***Capítulo 5. Discusión.***

## **Capítulo 5. Discusión.**

### **5.1. Vida Productiva**

El promedio de vida productiva (VP) para registros no censurados en este trabajo fue de 1058,32 días (2 años 10 meses aproximadamente). Este resultado es cercano al obtenido por Veerkamp et al. (1999), quienes observaron que cerca de la mitad de la población de vacas lecheras estudiada fue descartada antes de los tres años de VP. Sin embargo, es superior al valor obtenido por Roxström et al. (2003), en vacas de la raza Blanca y Roja Sueca (SRB), y por Caraviello et al. (2004a, b) en vacas Holstein y Jersey. En ambos trabajos se reportaron valores promedio de vida productiva para observaciones no censuradas de 594 días, 722 días y 807 días, respectivamente para las tres razas, o 1,63, 1,98 y 2,21 años. No obstante, el promedio obtenido para Holstein Colombiano fue menor al observado en ganado de carne con valores entre 8 a 9 años (Tarrés et al., 2004; Menéndez-Buxadera y Dempfle, 1997; Díaz et al., 2002). Sin embargo y como indicó Mészáros (2010), los resultados obtenidos son altamente dependientes del número de observaciones censuradas y de la longitud del período de estudio. Además, es probable que la variación en la tasa de censura y en el tiempo de VP se deba a diferencias regionales del tamaño de hato, clima, alimentación y condiciones de manejo, entre otros (Roxström et al., 2003).

Asumir un modelo de riesgo proporcional para VP fue razonable en el conjunto de datos evaluado, dada la proporcionalidad observada a través de los diferentes grupos de edad al primer parto (ver figura 4.1.1.1). Este resultado es similar al obtenido tanto por Roxström et al. (2003) como por Sánchez et al. (2004), quienes observaron una disparidad en el inicio del periodo de análisis a través de estratos (número de partos y generación respectivamente). En dichos casos, los autores asumieron proporcionalidad en las funciones de riesgo. Específicamente, las diferencias observadas en el inicio del período, tanto en el presente trabajo como en los citados anteriormente, estaría relacionada con un bajo número de animales por estrato (Ducrocq et al, 2000).

#### **5.1.1. Ajuste Weibull para la función de riesgo base.**

A nivel computacional, asumir una distribución Weibull para la función de riesgo base es relevante dada la simplicidad de dicho modelo al momento de calcular el vector de primeras derivadas del logaritmo de la función de verosimilitud (Ducrocq et al., 2000). Para el conjunto de datos estudiado, se observaron cambios en la pendiente de la recta de regresión de  $\log[-\log(S_{KM}(t))]$  versus  $\log(\text{tiempo})$ , sugiriendo la conveniencia de incluir una variable tiempo-dependiente que contemple los cambios en el riesgo a través del tiempo. En general, la falta de ajuste completo de la recta de regresión puede estar relacionada con una alta tasa de descarte durante el periodo voluntario de espera, más notable en los primeros 20 días de dicho periodo. El resultado obtenido es similar al presentado por Roxström et al. (2003), quienes observaron cuatro intervalos de ajuste, dando lugar a la inclusión en el análisis de un efecto tiempo-dependiente de estado de lactancia. Al igual que estos autores, en el presente trabajo fue difícil decidir cuáles serían los límites para cada uno de los intervalos en la recta de regresión (ver figura 4.1.2.1). Sin embargo, se consideraron segmentos rectos que presentaban pendientes diferentes de un intervalo a otro, hecho que sugiere que el supuesto Weibull se mantenía dentro del intervalo. Con lo cual, y buscando un adecuado ajuste del modelo Weibull, fue necesario un cambio en los parámetros de la



distribución de un intervalo a otro, aspecto que fue tenido en cuenta al incluir la variable tiempo-dependiente días post-primer parto (DPP). Además, y si bien el resultado obtenido en este trabajo para VP fue similar al ajuste presentado por Yazdi et al. (2000), estos investigadores asumieron un modelo Weibull para todo el período de estudio aun teniendo falta de ajuste después del primer destete cuando hubo descarte intensivo. Esta elección se debió a la facilidad computacional mencionada previamente, obtenida al asumir una distribución Weibull para el riesgo base. No obstante, cuando se tiene falta de ajuste de la distribución Weibull en un conjunto de datos de escaso número de registros, se puede recurrir a un modelo semiparamétrico o modelo de Cox (Cox, 1972), tal como el presentado en Sánchez et al. (2004).

### **5.1.2. Modelo de efectos fijos.**

De acuerdo a los valores de  $-2\log(\text{verosimilitud})$  presentados en las tablas 4.1.3.1. y 4.1.3.2, los efectos fijos significativos en la VP de Holstein Colombiano fueron el número de partos (NP), el nivel de producción de leche (PL), el número de días post-primer parto (DPP) y la edad al primer parto (EPP), siendo los dos primeros los factores con mayor influencia en dicho carácter funcional. En el caso del efecto NP, el riesgo de descarte es similar en las primeras dos lactancias, para luego disminuir a partir del tercer parto y mantenerse aproximadamente constante hasta la séptima parición. Un resultado similar fue encontrado por Vukasinovic et al. (2001), quienes observaron aumento en el riesgo de descarte en vacas con 4 o más lactancias. Este resultado según los autores, está relacionado con el mayor riesgo de muerte o descarte por razones sanitarias o problemas reproductivos en vacas de mayor edad. Además, la tendencia similar observada en el riesgo de descarte para las dos primeras lactancias sería explicada por un criterio de selección diferente en vacas de primer parto (Keller y Allaire, 1987; Vukasinovic et al., 2001).

El nivel de PL tuvo un efecto marcado en la VP del Holstein Colombiano. Los resultados de ajustar el modelo [51] muestran que vacas con el más bajo nivel de producción tuvieron un riesgo mucho mayor que aquellas que se encontraban en niveles superiores. Este resultado es comparable con el obtenido por Ducrocq (1994), quien reportó un comportamiento similar para vacas de la raza Normando. De acuerdo con el autor, vacas de baja producción tuvieron de tres a cuatro veces más probabilidad de ser descartadas que aquellas hembras en niveles de PL superiores. Asimismo, las vacas de alta producción tuvieron menos riesgo de ser descartadas que aquellas de producción promedio, al igual que encontraron Ducrocq (1987) y Ducrocq et al. (1988a,b).

El efecto de EPP tuvo una menor influencia que los anteriores respecto de vida productiva. Sin embargo, y de acuerdo a la figura 4.1.3.1, se observó una tendencia curvilínea, resultado similar al obtenido por Vukasinovic et al. (2001) con vacas de la raza Simmental y Holstein. Al igual que el presente trabajo, los autores reportaron un mayor riesgo de descarte para vacas que presentaron su primer parto tanto temprana como tardíamente, siendo mayor el riesgo para este último grupo de hembras debido a su posible asociación con problemas de fertilidad. Sin embargo, el riesgo de descarte es mucho menor en vacas que presentaron su primer parto antes de los 24 meses en comparación con hembras que tuvieron su primera parición después de los 36 meses de edad. Ducrocq (1999b) observó que esto podía deberse a que las vacas referidas en primera instancia son menos penalizadas por su baja producción. Además, la tendencia observada estaría

indicando que el primer parto no debería ocurrir luego de los 36 meses de edad, momento a partir del cual aumenta el riesgo de descarte de la vaca por una posible asociación con la suba en los costos de reemplazo. En este punto es importante resaltar que el rango de la variable EPP dentro de la presente investigación (18 a 42 meses) permitió evaluar los efectos de partos precoces, es decir de antes de los 24 meses de edad, sobre la vida productiva, tal como se los practica en algunos hatos de Holstein Colombiano.

En cuanto al efecto que DPP tiene sobre VP, se observó que las vacas que se encontraban entre los 20 y 45 días post-primer parto mostraron 3,9 veces más riesgo de descarte que las hembras que continuaron su vida productiva después del periodo de espera voluntario asociado con la primera parición. Como se mencionó en el capítulo de revisión, el periodo DPP es no solo afectado por variables fisiológicas sino también por decisiones que toma el productor. Por lo tanto, el alto riesgo de descarte observado entre 20 y 45 días luego de la primera parición estaría mostrando la importancia de la aptitud de la vaca para recuperar su estado productivo y reproductivo en dicho período, afectando las decisiones de descarte por parte del productor.

### **5.1.3. Componentes de varianza y heredabilidad**

El estimador parámetro Weibull  $\rho$  para VP fue igual a 2,3, indicando un aumento en el riesgo de descarte a medida que transcurre la VP en las hembras. Este valor es superior a los reportados en la bibliografía, los cuales se encuentran entre 0,58 y 1,76 (Ducrocq, 1994; Durr et al., 1999; Vukasinovic et al., 2001; Roxström et al., 2003; Caraviello et al., 2004b). Asimismo, los valores estimados de varianza aditiva y de varianza de hato fueron 0,269 y 0,217, respectivamente. Si bien Roxström et al. (2003) reportaron un valor similar para varianza de hato-año-estación (0,23), este valor no es comparable con el obtenido en el presente trabajo para el efecto hato. Específicamente, se estimó la varianza de hato y no de hato-año-estación dado que Colombia, al estar ubicada en la zona ecuatorial, no cuenta con estaciones climáticas, dificultando ajustar dicha varianza por el efecto de estación. Además, las estimaciones de varianza genética no son estrictamente comparables porque en nuestro trabajo fueron realizadas con un modelo animal y no un modelo de padre, tal como en todas las anteriores referencias.

La heredabilidad estimada fue de 0,105, valor inferior al reportado por Vukasinovic et al. (2001, 0,18 - 0,20), pero más cercano al reportado por Roxström et al. (2003, 0,135 a 0,165). En general, las estimaciones de heredabilidad se encuentran en el rango de 0.02 a 0.20 para hatos lecheros (Vollema y Goren, 1998), y entre 0.11 y 0.14 (Snelling et al., 1995), 0.03–0.26 (Tanida et al., 1998), 0.05–0.15 (Martínez et al., 2004), y 0.14 (Rogers et al., 2004) en bovinos para carne. Sin embargo, es importante aclarar que los valores reportados en la bibliografía fueron obtenidos asumiendo, en la mayoría de los casos, tanto diferencias en las definiciones para los efectos aleatorios (varianza de hato-año-estación versus varianza de hato), como definiciones de heredabilidad diferentes a la empleada en la presente tesis (Mészáros et al., 2010). En estos trabajos, la heredabilidad fue estimada siguiendo la definición de heredabilidad equivalente para un modelo de padre, dada por Yazdi et al. (2002), e igual a:

$$h^2 = \frac{4\sigma_s^2}{1 + \sigma_s^2 + \sigma_h^2} \quad [59]$$

En un modelo de padre, y definiendo la varianza genética como  $\sigma_G^2 = 4\sigma_s^2$ , la expresión [59] es como sigue:

$$h^2 = \frac{\sigma_G^2}{1 + \frac{\sigma_G^2}{4} + \sigma_h^2} \quad [60]$$

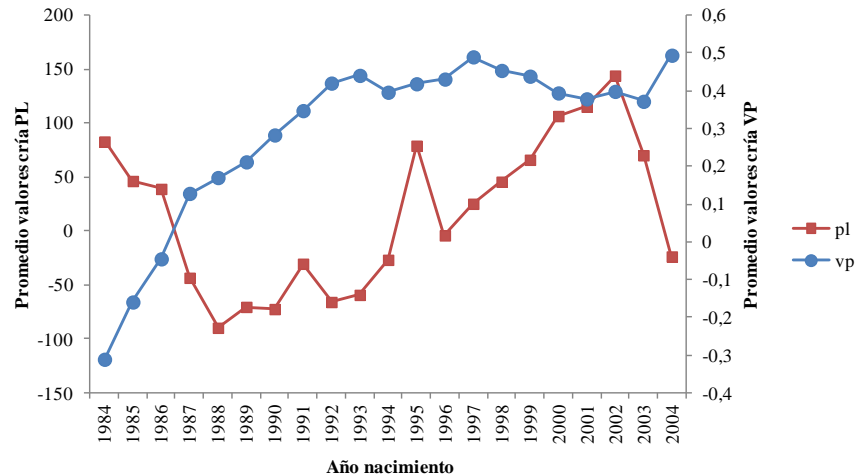
Ahora, si se asume un modelo animal, la heredabilidad teniendo en cuenta la expresión [60] estaría dada por:

$$h^2 = \frac{\sigma_G^2}{1 + \sigma_G^2 + \sigma_h^2} \quad [61]$$

Donde  $\sigma_G^2$  depende totalmente del efecto de la vaca. Al comparar las expresiones [61] y [53], debe notarse que está última definición tiene en cuenta el efecto de la proporción de individuos no censurados en la heredabilidad  $p$ , a diferencia de la expresión [61], la cual asume que todos los individuos han presentado el evento o falla, aspecto que puede conducir a diferencias en las estimaciones de heredabilidad.

#### 5.1.4. Tendencia genética

De acuerdo a los resultados, y teniendo en cuenta que en el modelo de supervivencia se evalúa el efecto del animal (o de su padre) sobre el riesgo relativo de descarte, la tendencia genética está relacionada con los valores de cría para riesgo de descarte. Con lo cual, y como se mencionó en el capítulo de resultados, un aumento en dichas estimaciones estaría mostrando una disminución en la VP de las hembras. Durante los años 1984 a 1992, el valor de cría predicho para riesgo de descarte aumentó constantemente dando lugar a una tendencia desfavorable para VP, hecho posiblemente debido al escaso número de vacas con varias pariciones en los primeros años. A partir de 1992 se observó cierta estabilidad en el comportamiento del promedio de los valores de cría predichos con una leve tendencia negativa (un aumento gradual en los últimos 7 años), resultados que concuerdan con los obtenidos por Ducrocq (1999b). Este autor halló una tendencia favorable de la VP en Holstein, Normando y Montbéliarde, en Francia. Sin embargo y de acuerdo con Essl (1998) la tendencia en la VP no está relacionada solo a causas genéticas sino también a factores ambientales. Así, Sölkner y Essl (1989) encontraron una influencia significativa sobre VP en vacas que habían sido previamente seleccionadas por producción de leche, en sistemas de estabulación. Al respecto, y de acuerdo con la literatura, la selección por alta producción de leche en la primera lactancia conduciría a un incremento en la vida productiva (Hoque y Hodges, 1980; Pape et al., 1982; Chauhan y Hayes, 1993; Jairath et al., 1995). Por lo tanto, se evaluó la tendencia genética para producción de leche para el mismo conjunto de datos evaluados para VP.



**Figura 5.1.4.1.** Tendencia genética para vida productiva y producción de leche.

La tendencia genética para producción de leche en la Figura 5.1.4.1 disminuyó en el período 1984-1988, para luego aumentar constantemente. Dicho retroceso coincide con el aumento en la tendencia para el riesgo de descarte. Es decir, en los primeros cuatro años VP y PL disminuyeron. Sin embargo, dado el cambio en la tendencia a partir de 1988 y con el fin de dilucidar la relación entre los valores de cría predichos para VP y PL, se estimó la correlación entre ambos caracteres igual a  $-0,3$ , empleando el enfoque de Calo et al. (1973). Dicha cifra implicaría que un mayor nivel de PL estaría relacionado con un menor riesgo de descarte y, por lo tanto, con un aumento en VP. Sin embargo, y de acuerdo a las observaciones de Essl (1984), este resultado es infrecuente en ganado lechero, donde un incremento sustancial en PL conduce a una disminución en el promedio de VP. Oltenacu y Broom (2010) indicaron que el aumento en PL suele relacionarse con la reducción en la fertilidad, en general, y en VP, en particular.

## 5.2. Intervalo primer-último servicio

Como se mencionó en el capítulo de revisión y de acuerdo con Jorjani (2005), el carácter intervalo entre el primer y último servicio (IPU) es un indicador de la fertilidad del ganado lechero. Sin embargo, Schneider et al. (2006) observaron que ajustar un modelo lineal a IPU puede conducir a resultados incorrectos debido a su distribución sesgada. Al igual que en la presente tesis, los autores emplearon un modelo de datos agrupados (Prentice y Gloeckler, 1978) para IPU, obteniendo mejor ajuste comparado con aquel obtenido con modelos lineales mixtos clásicos u otros modelos de supervivencia (modelo Weibull; modelo de Cox; Cox, 1972). Esto puede ser explicado por la capacidad de los modelos de datos agrupados para tener en cuenta la presencia de tiempos de falla similares (empates), situación que ocurre cuando las vacas quedan preñadas en la primera inseminación (39,38% en la presente tesis), y por tanto, su valor de IPU será el mismo.

Descriptivamente, el promedio de IPU antes de la discretización fue de 63,68 días. Este valor es superior al reportado en la literatura, donde el IPU promedio es de 23,4 a 44,8 días (Koeck et al., 2010; Dhaliwal, 1996; González-Recio et al., 2005; Ghiasi et al., 2011). Sin embargo, el promedio observado puede deberse a una baja tasa de detección de celos en los hatos lecheros Colombianos (37%; Ascholsstein, 2011). Este porcentaje se refleja en el

número de servicios por preñez, dado que el valor promedio observado en el presente trabajo (2,4 servicios por concepción) es superior al reportado por Dhaliwal (1996) (1,65 y 1,92 para vacas de baja y alta producción, respectivamente).

### **5.2.1. Modelo de datos agrupados**

De acuerdo con las tablas 4.2.1.1 y 4.2.1.2, los factores significativos en el análisis de IPU fueron el número de partos (NP) y el número de servicios (SC). La edad al primer parto (EPP) no mostró un efecto significativo en la duración de IPU, a diferencia de lo reportado por Maizón et al. (2004), quienes observaron que el riesgo de concepción después del primer servicio disminuyó cuando aumentó la EPP efecto que, según los autores, es un indicador del desempeño reproductivo de la vaca durante su vida productiva. Además, el retraso en el primer servicio es un factor importante que afecta el IPU, dado que por cada día en que se redujo el número de días al primer servicio, IPU disminuyó en 0,54 días (Britt, 1975).

En cuanto al efecto del número de partos sobre IPU, Petterson et al. (2006) observaron que NP no afectó significativamente a la variable respuesta. Contrariamente, en la presente investigación se encontró que el riesgo de concepción fue menor para vacas con más de tres pariciones. Este resultado es comparable con el observado por Maizón et al. (2004), quienes reportaron una disminución en el riesgo de concepción para vacas con más de dos partos. El resultado puede explicarse por la ocurrencia de problemas reproductivos y/o sanitarios en las hembras a partir de la segunda o tercera lactancia, factores que fueron evaluados por los autores. En dicho trabajo, se encontró que desordenes reproductivos como distocia, retención de placenta y metritis afectaron negativamente los caracteres de fertilidad, entre ellos, el intervalo entre el primero y el último servicio de las hembras lecheras.

Por su parte, el efecto SC significativamente afectó el carácter IPU. Esto podría relacionarse con que por cada servicio no exitoso en la vaca lechera, se deberá esperar cerca de 21 días para poder volver a inseminarla. Consecuentemente, para que IPU sea corto la tasa de detección de celos así como la efectividad del servicio deberían tener una eficacia del 100%, situación que no ocurre en los hatos lecheros Colombianos. Además, al tenerse pariciones continuas durante todo el año, la detección de celos sería uno de los factores determinantes de la eficiencia reproductiva del Holstein Colombiano, cuando se la compara con la de rodeos que se manejan con servicios estacionados, caso del ganado de carne.

### **5.2.2. Estimación de componentes de varianza y heredabilidad**

La varianza aditiva, la de hato y la heredabilidad para IPU fueron estimadas en 0,4, 1,00 y 0,10, respectivamente. En primer lugar, los componentes de varianza no son comparables con otros estudios dado que en trabajos previos donde se analizó el mismo carácter, se emplearon modelos de padre (Maizón et al., 2004; Schneider et al., 2006) en vez de modelos animales. Además, dichos trabajos han asumido para el efecto de hato distribuciones gamma en lugar de una distribución normal, como ha ocurrido en la presente investigación. De igual manera, y si bien la heredabilidad obtenida en la presente tesis para IPU es mayor a la reportada en la bibliografía (rango entre 0,004 a 0,048: Maizón et al.,

2004; Schneider et al., 2006; Roxström et al., 2001; Ghiasi et al., 2011), la situación descrita anteriormente en cuanto a tipo de modelo empleado y los supuestos distribucionales obscurece la comparación de los resultados. Sin embargo, aunque las heredabilidades estimadas fueron bajas, tanto con las metodologías revisadas como con la obtenida durante este trabajo, el resultado indica la presencia de variabilidad genética para IPU en Holstein Colombiano, sugiriendo que dicho carácter puede ser considerado en programas de selección de la raza (Philipsson, 1981).

### **5.2.3. Tendencia genética**

La figura 4.2.3.1 muestra que la tendencia genética para IPU en términos de riesgo de concepción fue ascendente. Esto implica que la selección en Holstein Colombiano ha favorecido la disminución del intervalo entre el primer y el último servicio, dado que los valores de cría estimados corresponden a riesgo de concepción. Como el carácter IPU está afectado por la detección de celos y por la eficiencia en el servicio, el resultado sugiere mejora en uno o ambos aspectos, conllevando a un mejor desempeño reproductivo de las hembras y a una mayor rentabilidad de la empresa agropecuaria. Al respecto, Mc Dowell (1972) estimó que por cada día que se extiende IPU, hubo pérdidas de hasta 1,7 dólares/vaca. Asimismo, González-Recio y Alenda (2005) destacaron la importancia económica de IPU teniendo en cuenta su correlación genética con otros indicadores reproductivos, como por ejemplo el intervalo entre partos. El resultado obtenido también puede estar relacionado con mejoras en el manejo de las hembras lecheras evaluadas, ya que un bajo nivel nutricional puede inducir falta de desarrollo ovárico y uterino y, como consecuencia de ello, producir retrasos en la manifestación del celo (León Velarde, 1979).

### **5.3. Días abiertos**

Philipsson (1981) y Jansen (1986) consideraron que la fertilidad en ganado lechero es el carácter más importante en términos económicos luego de la producción de leche. Sin embargo, el énfasis en seleccionar por PL, combinado a su antagónica correlación genética con fertilidad, ha resultado en una reducción de la capacidad de la vaca lechera para concebir y llevar a buen término una parición (Royal et al., 2000; Roxström et al. 2001). En el caso específico de los días abiertos (DA), Schneider et al. (2005) observaron que son resultantes de una combinación entre el retorno a la ciclicidad de la vaca después del parto, y de la expresión del celo y la habilidad para concebir. Sin embargo, Ghiasi et al. (2011), indicaron que DA puede hallarse afectado por decisiones de manejo tales como la longitud del periodo voluntario de espera, o la sincronización de celos que se realiza en algunos hatos.

Para el análisis de DA se emplearon por igual metodologías de modelos lineales mixtos (Liu et al., 2008) como de modelos de supervivencia (Lee et al., 1989; Schneider et al., 2005; Maizón et al., 2004). En el primer caso, Liu et al. (2008) concluyeron que aunque la información de inseminaciones incompletas era tratada como un valor desconocido en un modelo multicaracter, el modelo de fertilidad necesita ser mejorado para considerar información de servicios de vacas descartadas. Alternativamente, el modelo de supervivencia tiene la ventaja de expresar los resultados en la forma de probabilidades de concepción específicas a determinados tiempos, lo cual puede usarse en muchos tipos de

análisis económicos y de decisión. De esta manera, en la presente investigación, y al igual que el intervalo entre primer y último servicio, los DA fueron analizados mediante el ajuste de un modelo de supervivencia y más específicamente, el modelo de datos agrupados (Prentice y Gloeckler, 1978). Así para el conjunto de datos evaluado, el promedio de días abiertos fue de 138,72 días para observaciones no censuradas. Este valor es superior al reportado por Gredler (2007), Ghiasi et al. (2011) y Gonzalez-Recio et al. (2005), quienes observaron promedios para DA entre 105,6 y 117,67 días, resultado que puede estar relacionado con las diferencias de manejo después del parto que ocurren en los hatos lecheros de Colombia. Específicamente, factores nutricionales y reproductivos como sincronización y detección de celos, entre otros.

Lee et al. (1989) afirmaron que la ventaja del análisis de supervivencia reside en la factibilidad al estimar la variación en los DA en términos de la tasa de riesgo, que puede ser obtenida a partir de la función de riesgo. Asimismo, Schneider et al. (2005) mostraron mediante simulación, que el análisis de supervivencia fue mejor método de estimación que los modelos lineales, en la predicción del mérito genético de toros para DA.

### **5.3.1. Modelo de datos agrupados**

Los días abiertos estuvieron significativamente afectados por el número de partos (NP), el número de servicios por ciclo (SC) y la edad al primer parto (EPP). En primer lugar, vacas que tuvieron un mayor NP tuvieron un menor riesgo de concepción que aquellas vacas con escasas pariciones. Este resultado es similar al reportado por Maizón et al. (2004), quienes observaron que hembras en su segundo parto tuvieron un riesgo más alto de concepción que vacas de una o más de dos pariciones. En el caso del Holstein Colombiano, el menor riesgo de concepción para vacas de varios partos puede estar asociado con la mayor incidencia en dichas hembras de enfermedades antes o después de los 45 días postparto, tales como la distocia, la retención de placenta, y las metritis y/o mastitis, entre otras. De acuerdo con Lee et al. (1989) y Maizón et al. (2004), los problemas mencionados disminuyeron el riesgo de concepción en vacas de varios partos, dando lugar a un aumento en DA.

En el caso del efecto EPP, el riesgo a la concepción disminuyó cuando aumentó la EPP, resultado similar a lo reportado por Maizón et al. (2004), quienes obtuvieron mayor número de DA para vacas que tuvieron su primer parto luego de los 32 meses de edad. En general, la EPP está relacionada con la edad en que se produce el primer servicio, y dependerá del manejo y nivel nutricional de la hembra durante el crecimiento. Con lo cual, la EPP afecta significativamente la eficiencia reproductiva (León Velarde, 1979). Esta afirmación concuerda con la correlación genética entre EPP y DA estimada en 0,05 por Gill y Allaire (1976). Con lo cual, la importancia económica de DA en la rentabilidad del hato y su relación con la EPP deben ser contempladas, ya que un aumento en EPP de hembras lecheras conllevará a un posible aumento en los DA y, por lo tanto, a una mayor tasa de reposición. Asimismo, no es económica ni productivamente viable mantener vacas con amplios períodos de DA dado que esto generará a mediano y largo plazo, disminución en la producción diaria de leche y por tanto, en los ingresos del hato (Gill y Allaire, 1976; González-Recio et al., 2005). Por otra parte y de modo similar que en el análisis del carácter intervalo entre primer y último servicio, el efecto SC fue significativo debido,

posiblemente, al mayor número de servicios que se requieren en vacas con altos valores de DA.

### **5.3.2. Estimación de los componentes de varianza y la heredabilidad**

Para el carácter DA, las estimaciones de varianza aditiva, de hato y heredabilidad fueron 0,245, 0,292 y 0,086, respectivamente. Estos resultados, pero en particular la heredabilidad estimada, son cercanos a lo reportado en la bibliografía, donde se hallan valores de heredabilidad para DA en el rango de 0,04 a 0,076, cuando se ajustan modelos lineales mixtos (Gredler, 2007; Ghiasi et al., 2011). Maizón et al. (2004) reportaron un valor de heredabilidad para DA en la escala logarítmica de 0,012 ajustando modelos de supervivencia. Sin embargo, y al igual que para el intervalo entre el primer y último servicio, los resultados obtenidos en la presente investigación no son comparables con los valores encontrados en la bibliografía debido a la metodología estadística empleada (modelos lineales versus modelos de datos agrupados) y a los efectos aleatorios incluidos en el modelo (modelo de padre versus modelo animal). No obstante ello, el valor estimado para heredabilidad de DA muestra la existencia de variabilidad genética disponible para selección en Holstein Colombiano. De acuerdo a González-Recio et al. (2004), caracteres de intervalo como DA deben ser incluidos en índices de selección económicos, ya que en los esquemas de registros de producción de leche, se tiene disponibilidad de las fechas de parto y en algunos casos, de las fechas de servicio. Esto facilita la inclusión de los DA en los índices económicos, con el fin de analizar su relación con la rentabilidad del hato.

### **5.3.3. Tendencia genética.**

La tendencia genética para DA, expresada en términos de riesgo de concepción, mostró un comportamiento ascendente. Este resultado implicaría que en los últimos años, la selección en Holstein Colombiano ha dado lugar a una disminución en el número de DA, dado que los valores de cría de las hembras evaluadas para riesgo de concepción ha aumentado. Al respecto, y de acuerdo con Weaver y Goodger (1987), evaluar la tendencia del carácter DA permite analizar pérdidas de ingresos económicos asociadas con un mayor número de días en período seco de las vacas, así como con menor producción de leche y menos terneros por año. El resultado obtenido en la presente investigación difiere de lo reportado por Jorjani et al. (2007), quienes observaron que el mérito genético predicho promedio en poblaciones de Holstein Estadounidenses disminuyó cerca del 12% entre 1986 y 2000. De igual modo, González-Recio et al. (2004) concluyeron que la fertilidad en el ganado lechero español ha disminuido cerca del 10% en los últimos 14 años. Más específicamente, en esta última población los DA aumentaron aproximadamente en 15 días, debido al deterioro del desempeño reproductivo de las vacas.

En este punto debe mencionarse que, al igual que para la vida productiva funcional, se estimó la correlación simple (Calo et al., 1973) entre los valores de cría estimados para riesgo a la concepción y aquellos estimados para el nivel de producción de leche. En el presente conjunto de datos, la correlación simple fue estimada en 0,47, resultado que implicaría que a mayor producción de leche, habría mayor valor de riesgo de concepción y consecuentemente un menor número de DA. Si bien este resultado difiere con lo reportado en la bibliografía (Bell y Roberts, 2007; Dobson et al., 2007; Sheldon et al., 2008), Hodel et al. (1995) señalaron que es factible observar que vacas de menor producción de leche



tienen una menor probabilidad de ser inseminadas nuevamente, a diferencia de vacas de alta producción. Además, los autores afirmaron que este resultado puede relacionarse con el porcentaje de registros de vacas de primera lactancia en el conjunto de datos, así como del porcentaje de vacas inseminadas y que son descartadas.

## ***Capítulo 6. Conclusiones.***

## **Capítulo 6. Conclusiones**

El análisis de supervivencia fue una herramienta útil al momento de evaluar observaciones de tiempo a la ocurrencia de un evento, permitiendo la inclusión tanto de registros de vacas que no fueron descartadas al terminar el análisis, como información proveniente de hembras que no lograron una preñez durante el período de seguimiento. No obstante ello, el análisis de supervivencia no fue adecuado al momento de evaluar caracteres con tiempos repetidos a la ocurrencia del evento de interés. El enfoque de datos agrupados resultó apropiado al momento de trabajar con observaciones repetidas para caracteres como DA e IPU. Específicamente, el modelo de datos agrupados permitió tener en cuenta la presencia de empates, estimando conjuntamente la función de riesgo y los parámetros de regresión relacionados con los efectos fijos del modelo de análisis.

### **6.1. Vida productiva.**

La vida productiva (VP) fue significativamente afectada por los efectos fijos del número de partos, el nivel de producción de leche, el número de días post-primer parto y la edad al primer parto. El valor estimado del parámetro Weibull ( $\rho$ ) fue 2,3, indicando un aumento en el riesgo de descarte a medida que transcurre el tiempo. La heredabilidad de VP fue estimada en 0,105. La tendencia genética mostró un aumento inicial del riesgo de descarte durante los primeros ocho años evaluados, momento a partir del cual se mostró relativamente estable, con una leve disminución del riesgo de descarte en los últimos 7 años, dando lugar a un aumento gradual de la VP en la población. La correlación entre los valores de cría predichos entre VP y PL fue igual a  $-0,3$ . Consecuentemente, se puede señalar que los programas de mejoramiento genético en Holstein Colombiano han permitido mejorar, de forma gradual, el mérito genético de las hembras para VP funcional.

### **6.2. Intervalo entre primer y último servicio.**

La presencia de tiempos repetidos de ocurrencia a la preñez o empates, requirió la implementación de un modelo de datos agrupados, incluyendo una covariable tiempo-dependiente adicional que permitió tener en cuenta este aspecto. A diferencia de la edad al primer parto, el número de partos y el número de servicios por ciclo fueron efectos significativos en IPU. La heredabilidad del carácter IPU fue estimada en 0,101. La tendencia genética indicó que durante el periodo estudiado existió un comportamiento favorable de los valores de cría predichos en el tiempo.

### **6.3. Días abiertos**

Los efectos fijos que afectaron significativamente DA fueron el número de partos, la edad al primer parto y el número de servicios por ciclo. La inclusión de la variable tiempo-dependiente relacionada con el ajuste de datos agrupados fue significativa para DA. La heredabilidad estimada fue igual a 0,086. La correlación entre los valores de cría predichos de DA y PL fue igual a 0,47, indicando que cuando la PL aumenta, el riesgo de concepción también aumenta y, consecuentemente, DA disminuye en la población Holstein de Colombia.

## BIBLIOGRAFIA

- Aalen, O. O., Borgan, O. y Gjessing, H. K. 2008. Survival and event history analysis. Springer-Verlag, New York.
- Andersen, A.H., Korsgaard, I.R. y Jensen, J. 2000. Identifiability of parameters in and equivalence of animal and sire models for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits. Department of Theoretical Statistics, University of Aarhus. *Research reports*. 417:1–36.
- Asociación Holstein de Colombia – Asoholstein. 2011. Resumen Nacional de desempeño productivo y reproductivo.
- Bell, M.J. y Roberts, D.J. 2007. The impact of uterine infection on a dairy cow's performance. *Theriogenology*. 68:1074–1079.
- Boichard, D. 1990. Estimation of the economic value of conception rate in dairy cattle. *Livest. Prod. Sci.* 24:187–204.
- Bourdon, R.M. y Brinks, J.S. 1983. Calving date versus calving interval as a reproductive measure in beef cattle. *J. Anim. sci.* 57:1412–1417.
- Calo, L. L., McDowell, R. E. VanVleck, L. D. y Miller, P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J. Anim. Sci.* 37:676–682.
- Caraviello, D.Z., Weigel, K.A. y Gianola, D. 2004a. Comparison between a Weibull proportional hazards model and a linear model for predicting the genetic merit of US Jersey sires for daughter longevity. *J. Dairy Sci.* 87: 1469-1476.
- Caraviello, D.Z., Weigel, K.A. y Gianola, D. 2004b. Prediction of longevity breeding values for US Holstein sires using survival analysis methodology. *J. dairy science*. 87:3518–25.
- Chauhan, V.P.S. y Hayes, J.F., 1993. Relationships of first lactation yields with lifetime performance traits in Holstein cows. *J. Anim. Breed. Genet.* 110, 264–267.
- Cox, D.R. 1972. Regression models and life-tables (with discussion). *J. Royal Stat. Soc Series B.* 34:187–220.
- Cox, D.R. y Oakes D. 1984. Analysis of survival data. Chapman & Hall. London, UK.
- Damgaard, L.H., y Korsgaard, I.R. 2006a. A bivariate quantitative genetic model for a linear Gaussian trait and a survival. *Gen. sel. evol.* 38:45–64.
- Damgaard, L.H., y Korsgaard, I.R. 2006b. A bivariate quantitative genetic model for a threshold trait and a survival trait. *Gen. sel. evol.* 38:565–581.
- Dempster, A.P., Laird, N.M. y Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Stat Soc. Series B.* 39: 1-38.

- Dhaliwal, G.S., Murray, R.D. y Dobson, H., 1996. Effects of milk yield, and calving to first service interval, in determining herd fertility in dairy cows. *Anim. Reprod. Sci.* 41 (2):109–117.
- Díaz, C., Chirinos, Z., Moreno, A. y Carabaño, M. J. 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. *Proceedings of the seventh world congress on genetics applied to livestock production, Montpellier, vol. 29*, pp. 697-700.
- Dirven, M. 2001. Dairy Clusters in Latin America in the Context of Globalization. *International Food and Agribusiness Management Review*. 2(3/4): 301–313.
- Dobson, H., Smith, R., Royal, M., Knight, C. y Sheldon, I. 2007. The High-producing Dairy Cow and its Reproductive Performance. *Reproduction in Domestic Animals*. 42:17–23.
- Ducrocq, V. 1987. An analysis of length of productive life in dairy cattle. Ph.D. Diss., Cornell University, Ithaca, NY, USA.
- Ducrocq, V. 1994. Statistical analysis of length of productive life for dairy cows of the Normande breed. *J. Dairy Sci.* 77: 855-866.
- Ducrocq, V. 1999a. Extension of survival analysis models to discrete measures of longevity. Page 41 in Proc. Int. Workshop on Genetic Improvement of Functional Traits in Cattle, Longevity, Jouy-en-Josas, France. INTERBULL Bull. No. 21. Int. Bull Eval. Serv., Uppsala, Sweden.
- Ducrocq, V. 1999b. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. *Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity*. Jouy-en-Josas, France. INTERBULL Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden. pp. 60-70.
- Ducrocq, V. 2001. Survival analysis applied to animal breeding and epidemiology. Station de Génétique Quantitative et Appliquée. Institut National de la Recherche Agronomique. Jouy-en-Josas, France.
- Ducrocq, V. y Casella, G. 1996. A bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28:505-529.
- Ducrocq, V. y Sölkner, J. 1994. The Survival Kit – a Fortran package for the analysis of survival data. *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* Armidale, Australia. 27: 447-448.
- Ducrocq, V. y Sölkner, J. 1998. Implementation of a routine breeding value evaluation for longevity of dairy cows using survival Analysis techniques. *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* Armidale, Australia. 23: 359-362.

- Ducrocq, V., Besbes, B. y Protais, M. 2000. Genetic improvement of laying hens viability using survival analysis. *Genet. Sel. Evol.* 32:23-40
- Ducrocq, V., Sölkner, J y Mészáros, G. 2010. Survival Kit v6 - A software package for survival analysis. *9th World Congress on Genetics Applied to Livestock Production*, August 1-6. Leipzig.
- Ducrocq, V., Quaas, R. L., Pollak, E. J. y Casella, G. 1988a. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* 71: 3061-3070.
- Ducrocq, V., Quaas, R. L., Pollak, E. J. y Casella, G. 1988b. Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation. *J. Dairy Sci.* 71: 3071-3079.
- Ducrocq, V., Sölkner, J y Mészáros, G. 2010. Survival Kit v6 - A software package for survival analysis. *9th World Congress on Genetics Applied to Livestock Production*, August 1-6. Leipzig.
- Dürr, J. W., Monardes, H. G. y Cue, R.I. 1999. Genetic analysis of herd life in Quebec Holsteins using Weibull model. *J. Dairy Sci.* 82:2503–2513.
- Dziuk, P. J. y Bellows, R. A. 1983. Management of reproduction of beef caule. sheep and pigs. *J. Anim. Sci.* 57(Suppl. 2):355.
- Egger-Danner, C. 1993. Zuchtwertschätzung für Merkmale der Langlebigkeit beim Rind mit Methoden der Lebensdaueranalyse. Ph.D. Diss., Univ. of Agricultural Sciences. Vienna, Austria.
- Essl, A. 1998. Longevity in dairy cattle breeding : a review. 57:79–89.
- Everett, R.W., Armstrong, D.V. y Boyd, L.J. 1966. Genetic relationship between production and breeding efficiency. *J. Dairy Sci.* 49, 879-886.
- Famula, T.R. 1981. Exponential Stayability Model with Censoring and Covariates. *J. Dairy Science.* 64:538–545.
- Federación Colombiana de Ganaderos - Fedegan. 2012. Censo final: predios y población bovina por sexo y categorías de edad.
- Freeman, A. E. 1984. Secondary traits: Sire evaluation and the reproductive complex. *J. Dairy Sci.* 67:449–458.
- Ghiasi, H., Pakdel, A., Nejati – Javaremi, A., Mehrabani – Yeganeh, H., Honarvar, M., Gonzalez-Recio, O., Jesus Carabaño, M y Alenda, R. 2011. Genetic variance components for female fertility in Iranian Holstein cows. *Livest. Sci.* 139(3): 277-280.
- Gill, G.S. y Allaire, F.R. 1976. Relationship of Age at First Calving, Days Open, Days Dry, and Herdlife to a Profit function for Dairy Cattle. *J. Dairy Sci.* 59:1131–1139.

- Golden, B.L., Garrick, D.J., Newman, S., Enns, R.M., Collins, F. y Zealand, N. 2000. Economically relevant traits: A framework for the next generation of EPDs. *Proc. 32nd Beef Improv. Fed. Annu. Res. Symp. and Mtg.* Wichita, KS.
- González-Recio, O., y Alenda, R. 2005. Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. *J. Dairy Sci.* 88:3282–3289.
- González-Recio, O., Pérez-Cabal, M. y Alenda, R. 2004. Economic value of female fertility and its relationship with profit in Spanish dairy cattle. *J. Dairy Sci.* 87:3053–61.
- González-Recio, O., Chang, Y.M., Gianola, D. y Weigel, K.A. 2005. Number of inseminations to conception in Holstein Cows using censored records and time-dependent covariates. *J. Dairy Sci.* 88:3655–3662.
- Gredler, B., Fuerst, C., y Sölkner, J. 2007. Analysis of new fertility traits for the joint genetic evaluation in Austria and Germany. *Interbull Bull.* 37:152–155.
- Groen, A.F., Steine, T., Colleau, J.J., Pedersen, J. Pribyl, J. y Reinsch, N. 1997. Economic values in dairy cattle breeding, with special reference to functional traits. Report of an EAAP-working group. *Livestock Prod. Sci.* 49:1–21.
- Harbers, A. 1999. Durability breeding value in the Netherlands and the impact on sire rankings. *Proc. of the 1999 Interbull Meeting Zurich, Switzerland.* INTERBULL Bull No 22. Int. Bull. Eval. Ser., Uppsala, Sweden, pp. 126.
- Hare, E., Norman, H.D. y Wright, J.R. 2006. Trends in calving ages and calving intervals for dairy cattle breeds in the United States. *J. Dairy Sci.* 89:365–370.
- Harris, B.L., Freeman, E. y Metzger, E. 1992. Analysis of Herd Life in Guernsey Dairy Cattle. *J. Dairy Sci.* 75:2008–2016.
- Heins, B.J., Hansen, L.B. y Seykora, A.J. 2006. Production of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *J. Dairy Sci.* 89:2799–2804.
- Henderson, C.R. 1975. Best Linear Unbiased Estimation and Prediction under a selection model. *Biometrics.* 31: 423-447
- Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph. Guelph, Ontario, Canada.
- Henderson, C. R. y Quaas, R.L. 1976. Multiple trait evaluation using relatives records. *J. Anim. Sci.* 43:1188.
- Hermas, S. A., Young, C. W. y Rust, J. W. 1987. Genetic relationships and additive genetic variation of productive and reproductive traits in Guernsey dairy cattle. *J. Dairy Sci.* 70:1252–1257

- Hodel, F., Moll, J. y Kuenzi, N. 1995. Analysis of fertility in Swiss Simmental cattle - Genetic and environmental effects on female fertility. *Livest. Prod. Sci.* 41:95–103.
- Holmann, F., Rivas, L., Carulla, J., Rivera, B., Giraldo, L.A., Guzman, S., Martinez, M., Medina, A. y Farrow, A. 2003. Evolution of Milk Production Systems in Tropical Latin America and its interrelationship with Markets : An Analysis of the Colombian Case. 15.
- Holmann F, Rivas L, Carulla J, Rivera B, Giraldo LA, Guzmán S, Martínez M, Medina A, Farrow A. 2006. Producción de leche y su relación con los mercados; caso colombiano. Centro Internacional de Agricultura Tropical. X Seminario de Pastos y Forrajes. 149-156.
- Hoque, M. y Hodges, J., 1980. Genetic and phenotypic parameters of lifetime production traits in Holstein cows. *J. Dairy Sci.* 63,1900–1910.
- Hou, Y., Madsen, P., Laboriau, R. Zhang, Y., Lund, M.S. y Su, G. 2009. Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios. *J. Dairy Sci.* 92:1229–1239
- Hougaard, P. 1986a. A class of multivariate failure time distributions. *Biometrika.* 73: 671-678.
- Hougaard, P. 1986b. Survival models for heterogeneous populations derived from stable distributions. *Biometrika.* 73: 387-396.
- Jairath, L.K., Hayes, J.F. y Cue, R.I. 1995. Correlation between first lactation and lifetime performance traits of Canadian Holsteins. *J. Dairy Sci.* 78, 438–448.
- Jamrozik, J., Fatehi, J., y Schaeffer, L. R. 2008. Comparison of models for genetic evaluation of survival traits in dairy cattle: a simulation study. *J. Anim. Breed. Genet.* 124(1):3-11.
- Jamrozik, J., McGrath, S., Kemp, R. y Miller, S.P. 2012. Genetic analysis of female fertility traits in Canadian Simmentals. *Livest. Sci.* 150:302–309.
- Jansen, J. 1985. Genetic aspects of fertility in dairy cattle based on analysis of A.I. data—a review with emphasis on areas for further research. *Livest. Prod. Sci.* 12:1–12.
- Janson, L. 1980. Studies on fertility traits in Swedish dairy cattle. II: Genetic parameters. *Acta Agric. Scand.* 30:427–436.
- Jenkins, S.P. 2004. *Survival Analysis*. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK. Downloadable from <http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf>
- Jorjani, H. 2005. Preliminary report of Interbull pilot study for female fertility traits in Holstein populations. *Interbull Bulletin.* 33: 34-44.



- Jorjani, H. 2006. International genetic evaluation for female fertility traits. *Interbull Bulletin*. 34: 57-64.
- Jorjani, H., Jakobsen, J.H., Forabosco, F., Hjerpe, E. y Fikse, W.F. 2007. An international perspective on breeding for robustness in dairy cattle. 58th EAAP meeting, Dublin, Irlanda.
- Kalbfleisch, J.D. y Prentice, R. L. 1973. Marginal likelihoods based on Cox's regression and life model. *Biometrika* 60: 267-278.
- Kalbfleisch, J.D. y Prentice, R.L. 1980. The statistical analysis of failure time data. Wiley, New York. USA.
- Kalbfleisch, J. D. y Prentice, R. L. 2002. The Statistical Analysis of Failure Time Data. 2d ed. New York: John Wiley & Sons. Kay, R. 1986.
- Kaplan, E. L. y Meier, P. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53: 457-481.
- Keller, D.S. y Allaire, F.R., 1987. Relationships of first lactation milk and type traits to cow survival and sire index for discounted total milk. *J. Dairy Sci.* 70: 2116–2126.
- Klein, J. P. 1992. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*. 48: 795-806
- Klein, J.P. y Moeschberger, M. L. 2003. Survival analysis: techniques for censored and truncated data. 2nd ed. New York. Springer-Verlag.
- Klein, J. P., Moeschberger, M., Li, Y.H. y Wang, S. T. 1992. Estimation random effects in the Framingham heart study. En: *Survival Analysis: State of the Art*. Kluwer Academic Publishers, 99-120.
- Kleinbaum, D.G. 1996. Survival analysis, a self-learning text. New York. Springer-Verlag. 324p.
- Koeck, A., Egger-Danner, C., Fuerst, C., Obritzhauser, W. y Fuerst-Waltl, B. 2010. Genetic analysis of reproductive disorders and their relationship to fertility and milk yield in Austrian Fleckveih dual purpose cows. *J Dairy Sci.* 93:2185-2194.
- Korsgaard, I. R. 1996. Genetic analysis of survival data - a Gibbs sampling approach. International Workshop on Genetic Improvement of Functional Traits in Cattle. Faculte universitaire des sciences agronomiques. Gembloux, Belgium. Jan 21-23.
- Korsgaard, I. R., Madsen, P. y Jensen, J. 1998. Bayesian inference in the semi-parametric log normal frailty model using Gibbs sampling, *Genet. Sel. Evol.* 30: 241-256.
- Lawless, J.F. 1982. Statistical models and methods for lifetime data. New York: Wiley.
- Lee, L., Ferguson, J.D. y Galligan, D.T. 1989. Effect of disease on days open assessed by survival analysis. *J. Dairy Sci.* 72:1020–1026.

- León Velarde, C.U.1979. Fisiología y eficiencia reproductiva del ganado vacuno. Curso Metodología de Investigación y Generación de Sistemas de Producción Animal. CATIE. Turrialba, Costa Rica.
- Liu, Z., Jaitner, J. Reinhardt, F., Pasman, E., Rensing, S. y Reents, R. 2008. Genetic evaluation of fertility traits of dairy cattle using a multiple-trait animal model. *J. Dairy Sci.* 91:4333–4343
- Louca, A. & Legates, J.E. (1968) Production losses in cattle due to days open. /. *Dairy Sci.* 51, 573-583.
- McDowell, R.E. 1972. Improvement of Livestock production in Warm Climates. W.H.Freeman y Co., San Francisco, CA.
- Maizon, D.O., Oltenacu, P.A., Gröhn, Y.T., Strawderman, R.L. y Emanuelson, U. 2004. Effects of disease on reproductive performance in Swedish Red and White dairy cattle. *Prev. Vet. Med.* 66:113-126
- Mark, T. 2004. Applied genetic evaluations for production and functional traits in dairy cattle. *J. Dairy Sci.* 87:2641–52.
- Martinez, G.E., Koch, R.M., Cundiff, L.V., Gregory, K.E. y, Van Vleck, L.D., 2004. Number of calves born, number of calves weaned, and cumulative weaning weight as measures of lifetime production for Hereford cows. *J. Anim. Sci.* 82, 1903–1911.
- McAllister, A. J. 2002. Is crossbreeding the answer to questions of dairy breed utilization?. *J. Dairy Sci.* 85:2352–2357.
- Menéndez-Buxadera, A. y Dempfle, L. 1997. Genetic and environmental factors affecting some reproductive traits of Holstein cows in Cuba. *Genet Sel Evol.* 29: 469-482.
- Mészáros, G., Pálos, J., Ducrocq, V. y Sölkner, J. 2010. Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. *Genet. Sel. Evol.* 42:13
- Meyer, K., Hammond, K., Parnell, P.F., Mackinnon, M. J. y Sivarajasingam, S. 1990. Estimates of heritability and repeatability for reproductive traits in Australian beef cattle. *Livest. Prod. Sci.* 5:15–30.
- Miller, R. 1981. Survival analysis. Wiley, New York, USA.
- Ministerio de Agricultura y Desarrollo Rural - Federación Colombiana de Ganaderos - Fedegan. 2012. Coyuntura lactea - Contexto leche producción nacional.
- Nebel, R. L., and M. L. McGilliard. 1993. Interactions of high milk yield and reproductive performance in dairy cows. *J. Dairy Sci.* 76:3257.
- Oltenacu, P.A., y Broom, D.M. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal Welfare.* 19(S):39–49.

- Pape, H.Ch., Claus, J., Klam, E., 1982. Parameterschätzung zur Lebensleistung von Milchkuhen. *Züchtungskunde* 54, 257–266.
- Pasman, E. y Reinhardt, F. 1999. Genetic evaluation for length of productive life of Holstein cattle in Germany. *Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity*. Jouy-en-Josas, France. Interbull Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden. pp. 55
- Petersson, K-J., Strandberg, E., Gustafsson, H. y Berglund, B. 2006. Environmental effects on progesterone profile measures of dairy cow fertility. *Anim. Reprod. Sci.* 91: 201-214
- Philipsson, J., 1981. Genetic aspects of female fertility in dairy cattle. *Livest. Prod. Sci.* 8, 307–319.
- Prentice, R.L. y Gloeckler, L.A. 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*. 34: 57-67
- Prentice, R.L., y Kalbfleisch, J.D. 1979. Hazard Rate Models with Covariates. *Biometrics*. 35:25–39.
- Renkema, J.A. y Stelwagen, J., 1979. Economic evaluation of replacement rates in dairy herds. I. Reduction of replacement rates through improved health. *Livest. Prod. Sci.*, 6: 15--27.
- Robertson, A. 1977. The effect of selection on the estimation of genetic parameters. *Z. Tierz. Zuechtungsbiol.* 94:131-135.
- Rogers, P.L., Gaskins, C.T., Johnson, K.A. y MacNeil, M.D., 2004. Evaluating longevity of composite beef females using survival analysis techniques. *J. Anim. Sci.* 82, 860–866.
- Roxström, A., Ducrocq, V. y Strandberg, E. 2003. Survival analysis of longevity in dairy cattle on a selection basis. *Genet. Sel. Evol.* 35:305–318
- Sánchez, J. P., Peiró, R., Silvestre, M. A. y Baselga, M. 2004. Analysis of factors influencing longevity of rabbit does. *Livest. Prod. Sci.* 90:227–234
- Schaeffer, L. R. y Henderson, C. R. 1972. Effects of days dry and days open on Holstein milk production. *J. Dairy Sci.* 55(1):107-112.
- Schneider, M. del P., Cantet, R.J.C. y Santos Cristal de Sivak, M. 2002. Análisis de supervivencia en la evaluación genética de vida productiva en rodeos lecheros: una introducción. *Revista Argentina de Producción Animal.* 22:127–139.
- Schneider, M. del P., Strandberg, E. Ducrocq, V. y Roth, A. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. *J. Dairy Sci.* 88:2253-2259.

- Schneider, M. del P., Strandberg, E., Ducrocq, V y Roth, A. 2006. Short communication: genetic evaluation of the interval from first to last insemination with survival analysis and linear models. *J. Dairy Sci.* 89:4903–4906.
- Sheldon, I.M., Williams, E.J., Miller, A.N.A., Nash, D.M. y Herath, S. 2008. Uterine diseases in cattle after parturition. *Vet. J. London England*. 176:115–121.
- Silva, H.M., Wilcox, C.J., Thatcher, W.W., Becker, R.B. y Morse, D. 1992. Factors affecting days open, gestation length, and calving interval in Florida dairy cattle. *J. Dairy Sci.* 75:288–293.
- Simianer, H., Solbu, H., Schaeffer, L.R., 1991. Estimated genetic correlations between disease and yield traits in dairy cattle. *J. Dairy Sci.* 74: 4359–4365.
- Smith, S. P. 1983, The extension of failure time analysis to problems of animal breeding. *Ph.D. thesis*. Cornell Univ., Ithaca, NY.
- Smith, P.J .2002 Analysis of Failure and Survival Data. Boca Raton. Chapman and Hall/CRC; 2002.
- Smith, S.P., y Quaas, R.L. 1984. Productive lifespan of bull progeny groups: Failure time analysis. *J. Dairy Sci.* 67:2999–3007.
- Snelling, W.M., Golden, B.L. y Bourdon, R.M. 1995. Within-Herd Genetic Analyses of Stayability of Beef Females. *J. Anim. Sci.* 73:993–1001.
- Sölkner, J. y Essl, A. 1989. Einfluß verschiedener Formen der Anbindehaltung auf die Nutzungsdauer von Kühen Züchtungskunde. 62:222–233.
- Swalve, H.H., Topf, C. y Langholz, H. 1992. Estimation of genetic parameters for differently recorded fertility data in dairy cattle. *J. Anim. Breed. Gen.* 109:241–251.
- Tanida, H., Hohenboken, W.D. y DeNise, S.K., 1998. Genetic aspects of longevity in Angus and Hereford cows. *J. Anim. Sci.* 66, 640–647.
- Tarrés, J., Puig, P., Ducrocq, V y Piedrafita, J. 2004. Factors influencing length of productive life and replacement rates in the Bruna dels Pirineus beef breed. *Animal Science*. 78: 13-22
- Turnbull, B. 1976. The empirical distribution function with arbitrarily grouped, censored, and truncated data. *J. Royal Stat. Soc. Series B.* 38:290-295.
- Van Arendonk, J. A. M. 1991. Use of profit equations to determine relative economic value of dairy cattle herd life and production from field data. *J. Dairy Sci.* 74:1101.
- VanRaden, P.M., y Klaaskate, E.J.H. 1993. Genetic Evaluation of Length of Productive Life Including Predicted Longevity of Live Cows. *J. Dairy Sci.* 76:2758–2764.
- VanRaden, P. M., y Sanders, A. H. 2003. Economic merit of crossbred and purebred US dairy cattle. *J. Dairy Sci.* 86:1036–1044.

- Vaupel, J.W., Manton, K.G. y Stallard, E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 16:439–454.
- Veerkamp, R. F., Brotherstone, S. y Meuwissen, T. H. E. 1999. Survival analysis using random regression models. *Proceedings of the GIFT workshop on longevity*. Jouy-en Jossas, France. INTERBULL Bulletin 21: 36-40.
- Vollema, A.R. y Goren, A.F. 1998. A comparison of breeding value predictors for longevity using linear model and survival analysis. *J. Dairy Sci.* 81, 3315–3320.
- Vukasinovic, N., Moll, J. y Casanova, L. 2001. Implementation of a routing genetic evaluation for longevity based on survival analysis techniques in dairy cattle populations in Switzerland. *J. Dairy Sci.* 84:2073–2080.
- Warren, M.E. 1984. Biological targets and their effects on herd economics. *Dairy cow fertility*. British Veterinary Association Editorial Services. London.
- Weaver, L.D. y Goodger, W.J. 1987. Design and economic evaluation of dairy reproductive health programs for large dairy herds. *Part 1. Compend Contin Educ Pract Vet.* F297-F309.
- Weller, J. I. y Ron, M. 1992. Genetic analysis of fertility traits in Israeli Holsteins by linear and threshold models. *J. Dairy Sci.* 75: 2541-2548.
- Whitmore, H. L., Tyler, W. J. y Casida, L. E. 1974. Effects of early postpartum breeding in dairy cattle. *J. Anim. Sci.* 38:339.
- Yazdi, M., Rydhmer, L., Ringmar-Cederberg, E., Lundeheim, N. y Johansson, K. 2000. Genetic study of longevity in Swedish Landrace sows. *Livest. Prod. Sci.* 63:255–264.
- Yazdi, M. H., Visscher, P. M., Ducrocq, V. y Thompson, R. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *J. Dairy Sci.* 85:1563-1577.